Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structural Evolution of the Eukaryotic Ribosome

Andreas Markus Anger

aus München

2013

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung
vom 28. November 2011 von Herrn Prof. Dr. Roland Beckmann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 17.10.2013

_____

Andreas Anger

Dissertation eingereicht am 21.10.2013

1. Gutachter Prof. Dr. Roland Beckmann
2. Gutachter Dr. Daniel N. Wilson

Mündliche Prüfung am 20.11.2013

This cumulative thesis is based on the following peer-reviewed original publications (reprints are attached):

Paper 1:

Armache, J.-P.*, Anger, A. M.*, Márquez, V., Franckenberg, S., Fröhlich, T., Villa, E., Berninghausen, O., Thomm, M., Arnold, G. J., Beckmann, R., and Wilson, D. N. (2013). **Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution.** *Nucleic Acids Res.*, 41(2), 1284-1293.
* These authors contributed equally to this work.

Paper 2:

Armache, J.-P.*, Jarasch, A.*, Anger, A. M.*, Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., Marquez, V., Mielke, T., Thomm, M., Berninghausen, O., Beatrix, B., Söding, J., Westhof, E., Wilson, D. N., and Beckmann, R. (2010). **Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution.**
Proc. Natl. Acad. Sci. USA, *107*(46), 19748–19753.
* These authors contributed equally to this work.

Paper 3:

Armache, J.-P.*, Jarasch, A.*, Anger, A. M.*, Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., Marquez, V., Mielke, T., Thomm, M., Berninghausen, O., Beatrix, B., Söding, J., Westhof, E., Wilson, D. N., and Beckmann, R. (2010). **Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome.**
Proc. Natl. Acad. Sci. USA, *107*(46), 19754–19759.
* These authors contributed equally to this work.

Paper 4:

Anger, A. M.*, Armache, J.-P.*, Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., and Beckmann, R. (2013). **Structures of the human and *Drosophila* 80S ribosome.**
Nature, 497(7447), 80-85.
* These authors contributed equally to this work.

Parts of this thesis have been presented at international conferences:

2013 **Structures of the human and *Drosophila* ribosome.** (Poster)
Andreas M. Anger (presenting author), Jean-Paul Armache, Otto Berninghausen, Michael Habeck, Marion Subklewe, Daniel N. Wilson and Roland Beckmann

Ribosomes 2013 Conference
July 9-12, 2013, Napa Valley California, USA

2012 **Molecular architecture of human and *Drosophila* 80S ribosomes.** (Talk)
Andreas M. Anger (presenting author), Jean-Paul Armache, Otto Berninghausen, Charlotte Ungewickell, Marion Subklewe, Daniel N. Wilson and Roland Beckmann

13th Cold Spring Harbor Laboratory (CSHL) Conference on Translational Control
September 4-8, 2012, CSHL, Cold Spring Harbor New York, USA

2011 **Cryo-EM structure of the *Drosophila melanogaster* 80S ribosome.** (Poster)
Andreas M. Anger (presenting author), Jean-Paul Armache, Elizabeth Villa, Charlotte Ungewickell, Otto Berninghausen and Roland Beckmann

EMBO Conference on Protein Synthesis and Translational Control
September 7-11, 2011, EMBL Advanced Training Center, Heidelberg, Germany

2010 **The structure of a translating eukaryotic ribosome at 5.5 Å.** (Talk)
Jean-Paul Armache, Alexander Jarasch, Andreas M. Anger (presenting author), Elizabeth Villa, Thomas Becker, Shashi Bhushan, Fabrice Jossinet, Michael Habeck, Gülcin Dindar, Sibylle Franckenberg, Viter Márquez, Thorsten Mielke, Michael Thomm, Otto Berninghausen, Birgitta Beatrix, Johannes Söding, Eric Westhof, Daniel N. Wilson and Roland Beckmann

12th Cold Spring Harbor Laboratory (CSHL) Conference on Translational Control
September 13-17, 2010, CSHL, Cold Spring Harbor New York, USA

Table of Contents

Acknowledgements

First of all I would like to thank Roland Beckmann for being a great supervisor and mentor. I appreciate your advices, your constant support and the huge freedom you gave me to pursue my ideas. Experiencing your enthusiasm about science left an impression on me and I am very grateful for the opportunity to work in your lab.

Jean-Paul Armache, my blood brother. Thank you for sharing so many projects with me, the tremendous amount of time you invested in each of them and for your help regarding computer issues. It was great to work with you! Many thanks go to Daniel Wilson for all discussions, advices and constructive criticism, as well as for being the driving force behind our manuscripts.

In the following I want to thank a couple of people I had the privilege to work with over the last years: Sibylle Franckenberg and Viter Márquez for the collaboration on the archaea project; Viter, I'm also very grateful for your help during *Drosophila* embryo collections and mass spectrometry analysis; Alexander Jarasch and Elizabeth Villa for the collaboration on the wheat germ and yeast ribosome modeling project. It has been a rough ride and I'm glad we brought it to a success; Otto Berninghausen and Charlotte Ungewickell for handling the microscopes, collecting data and their support with everything concerning cryo-EM; Thomas Becker for teaching me how to process my datasets and Andreas Hauser for managing the cluster; Birgitta Beatrix for giving me useful advice regarding my experiments; Fabrice Jossinet for constantly improving his RNA modeling software and all the help when I ran into problems using it; Joanna Musial, Andrea Gilmozzi and Heidi Sieber for keeping the lab up and running; Ingegerd Walz for managing all non-lab tasks; Klaus Förstemann for introducing me to the *Drosophila* world and his kind support whenever I needed the cell culture. Peter Becker for access to his fly facility and Ignasi Forné for his support during mass spectrometric analysis of the human ribosome. I would like to thank my students Sarah Matheisl, Susanne Ciniawsky, Laura Kremer and Sebastian Johansson for their motivation and contributions. I really enjoyed working with you and wish you all the best for your future careers. I am also grateful to Monika Anger for the time she spent proofreading the human and *Drosophila* rRNA secondary structure diagrams. Thanks to all past and present members of the Beckmann and Wilson groups for the great and supportive atmosphere during my time here.

Above all, my biggest thanks go to Gülcin, for her love and support and to my parents, who were always there for me and made everything possible.

Summary

Ribosomes synthesize proteins following genetic information encoded in mRNA across all kingdoms of life. Despite the universal conservation of this process bacterial and eukaryotic ribosomes differ significantly in the complexity of their architecture and these structural differences are thought to reflect the more complex mechanisms of ribosome biosynthesis, translational initiation and regulation operating in the eukaryotic domain of life. Although crystal structures of bacterial ribosomes are available since more than a decade, high-resolution structures of eukaryotic ribosomes have only become available recently and are still limited to lower unicellular eukaryotes such as yeast.

Based on cryo-electron microscopy and single particle reconstruction this work reports molecular models of several eukaryotic ribosomes (yeast, wheat germ, fruit fly and human), covering a spectrum of organisms that includes representatives from lower and higher eukaryotes. The structures reveal eukaryote-specific rRNA and r-protein elements, their interactions with each other as well as with the universally conserved ribosome core. An intertwined architecture derived from coevolution of rRNA and r-proteins is found to be present in all eukaryotic ribosomes with some species-specific variations. In the human ribosome, however, it is further extended. Here two additional structural layers are observed, a well-ordered inner layer that is stabilized by unique RNA-RNA interactions, covered by a flexible RNA outer layer that forms tentacles protruding from the surface of the ribosome. The tentacles are likely to contribute to additional functionality of the ribosome in the context of the more complex mammalian cell. This work also presents the first complete molecular model of an archaeal 70S ribosome from *Pyrococcus furiosus*. The model illustrates that archaeal rRNAs of both subunits are chimeras of the corresponding bacterial and eukaryotic structures and uncovers a surprising promiscuity of r-proteins, with S24e and L8e being present on both ribosomal subunits. In addition, L8e and L14e exhibit intrasubunit promiscuity, each existing in two copies within the large subunit. The observation that the additional copies of L8e and L14e occupy positions where the related eukaryotic proteins S12e and L27e are located suggests that these eukaryotic r-proteins evolved through increased copy number and binding site promiscuity. Taken together, this dissertation gives insights into the evolution of the eukaryotic ribosome structure on both the RNA and protein level. The presented models provide the basis for more detailed structural, biochemical and genetic experiments, especially for the higher eukaryotes *Drosophila melanogaster* and human itself.

Contribution Report

This dissertation is based on work, which was conducted during my PhD research in the lab of Prof. Roland Beckmann from July 2008 to August 2013 at the Gene Center of the Ludwig-Maximilians University, Munich.

Paper 1 (Armache, Anger et al., 2013):

This paper presents the first complete molecular model of an archaeal 70S ribosome from *Pyrococcus furiosus* based on a 6.6 Å cryo-EM reconstruction. I built the *P. furiosus* rRNA model and performed the kink-turn analysis of the structure, which identified multiple binding sites for protein L8e within the rRNA. Moreover, I prepared all figures and contributed to writing of the manuscript.

Paper 2 (Armache, Jarasch et al., 2010a):

The publication reports first complete rRNA models of translating eukaryotic 80S ribosomes from *Triticum aestivum* and the yeast *Saccharomyces cerevisiae* based on cryo-EM reconstructions at 5.5 and 6.1 Å resolution, respectively. I build the rRNA models together with Alexander Jarasch and prepared all secondary structure diagrams (Figures 2A,B; 3A,B; 4A,C and Supplementary Figures S2 to S7). In addition, I contributed to the design of Figures 4 and 5, prepared Supplementary Figures S8 and S9, and participated in writing of the manuscript.

Paper 3 (Armache, Jarasch et al., 2010b):

This paper reports the localization of eukaryote specific ribosomal proteins in cryo-EM maps of *T. aestivum* and *S. cerevisiae* ribosomes. The models described in this publication complement the rRNA models from paper 3 to give a near complete molecular picture of the eukaryotic 80S ribosome. I designed Figure 4B and contributed to writing of the manuscript.

Paper 4 (Anger, Armache et al., 2013):

The publication reports first complete molecular models of the 80S ribosome from two distinct higher eukaryotic organisms, namely the fruit fly *Drosophila melanogaster* and human itself. I prepared embryo extracts from *D. melanogaster*, contributed blood for the isolation of peripheral blood mononuclear cells (PBMCs) and subsequently purified the human and *Drosophila* 80S ribosome samples from PBMCs and embryo extracts, respectively. I processed cryo-EM datasets and built the atomic models together with Jean-Paul Armache. Furthermore, I performed mass spectrometry analysis of the human ribosome sample with the help of Ignasi Forné. Finally, I prepared all figures and participated in writing of the manuscript.

Abbreviations

2D	two-dimensional
aa	amino acid
AAA	ATPase associated with diverse cellular activities
ADP	adenosine diphosphate
A-site	aminoacyl-tRNA site
ATPase	adenosine triphosphatase
A-tRNA	aminoacyl-tRNA
Be	beak
BF	b-factor
Bo	body domain
CP	central protuberance
cryo-EM	cryo-electron microscopy
C-terminus	carboxy-terminus
DC	decoding center
D. melanogaster	*Drosophila melanogaster*
DNA	desoxyribonucleic acid
rDNA	ribosomal DNA
D. radiodurans	*Deinococcus radiodurans*
eB#	eukaryote-specific intersubunit bridge #
E. coli	*Escherichia coli*
eEF	eukaryotic elongation factor
EF	elongation factor
eIF	eukaryotic initiation factor
eRF	eukaryotic release factor
ES	expansion segment
ES#L	expansion segment # (of the large subunit)
ES#S	expansion segment # (of the small subunit)
E-site	exit-tRNA site
E-tRNA	exit site tRNA
GTP	guanosine triphosphate
GTPase	guanosine triphosphatase
H	head domain
H#	RNA helix # (of the large subunit)
h#	RNA helix # (of the small subunit)
H. marismortui	*Haloarcula marismortui*
H. sapiens	*Homo sapiens*
ID	identifier
IF	initiation factor
KH	heterogeneous nuclear ribonucleoprotein K homology
KT	kink-turn
Lf	left foot
LSU	large subunit
M	methylation
MDa	megadalton
MS	mass spectrometry

M. thermoautotrophicus	*Methanobacterium thermoautotrophicus*
mTOR	mammalian target of rapamycin
NMR	nuclear magnetic resonance
nt(s)	nucleotide(s)
N-terminus	amino-terminus
PABP	poly-A binding protein
PAGE	polyacrylamide gel electrophoresis
PDB	protein data bank
P. furiosus	*Pyrococcus furiosus*
P_i	inorganic phosphate
PIC	pre-initiation complex
PKC	protein kinase C
ψ	pseudouridylation
P-site	peptidyl-tRNA site
Pt	platform
PTC	peptidyl transferase center
P-tRNA	peptidyl-tRNA
RAC	ribosome-associated complex
RACK1	receptor of activated C kinase 1
RF	release factor
Rf	right foot
RMSD	root mean square deviation
RNA	ribonucleic acid
mRNA	messenger RNA
rRNA	ribosomal RNA
tRNA	transfer RNA
RNase	ribonuclease
RPL	ribosomal protein of the large subunit
r-protein	ribosomal protein
RPS	ribosomal protein of the small subunit
RRF	ribosome release factor
S	sedimentation coefficient (Svedberg unit)
SB	stalk base
S. cerevisiae	*Saccharomyces cerevisiae*
SD	Shine-Dalgarno
Sh	shoulder
SH3	Src homology 3
snoRNP	small nucleolar ribonucleoprotein particle
snRNP	small nuclear ribonucleoprotein particle
Sp	spur
SSU	small subunit
SXL	sex-lethal
T. aestivum	*Triticum aestivum*
TC	ternary complex
TE	tunnel exit
T. kodakaraensis	*Thermococcus kodakaraensis*
T. thermophila	*Tetrahymena thermophila*
T. thermophilus	*Thermus thermophilus*
VR	variable region

1 Introduction

According to the central dogma of molecular biology genetic information flows from DNA via RNA to protein (Crick, 1970). Every organism is defined by the unique information content that is stored in its DNA. This information is propagated to the next generation via DNA replication and is utilized in the cell by means of transcription. During transcription, DNA-dependent RNA polymerases synthesize RNA molecules following the instructions of the DNA template. RNA molecules fulfill numerous roles in the cell. These include functions as information carriers, regulators of gene expression, structural scaffolds and enzymes. Messenger RNAs (mRNAs) encode information in the form of nucleotide triplets (codons) (Crick et al., 1961) and are templates for the translation of the genetic code into a chain of amino acids, called proteins. This last step in the flow of genetic information is carried out by ribosomes, cytosolic particles that were first described in 1955 by George Palade (Palade, 1955). Ribosomes decode mRNA with the help of transfer RNAs (tRNAs) that read codons one at a time and carry the corresponding amino acid. Proteins synthesized by the ribosome subsequently fold into their functional conformation with the help of molecular chaperones and participate in virtually every process in the cell.

1.1 The Ribosome: An Overview

Ribosomes are the universally conserved, macromolecular enzymes responsible for protein biosynthesis, the translation of genetic information from mRNA into polypeptides. The ribosome is composed of a small and large subunit (SSU and LSU, respectively), each built up from ribosomal RNA (rRNA) and ribosomal proteins (r-proteins). Both subunits have different roles during the translation process. The SSU harbors the decoding center (DC) where mRNA codon triplets are read, while the LSU contains the peptidyl transferase center (PTC), that catalyzes the linkage of amino acids to form proteins (Figure 1) (Schmeing and Ramakrishnan, 2009; Steitz, 2008; Voorhees and Ramakrishnan, 2013).

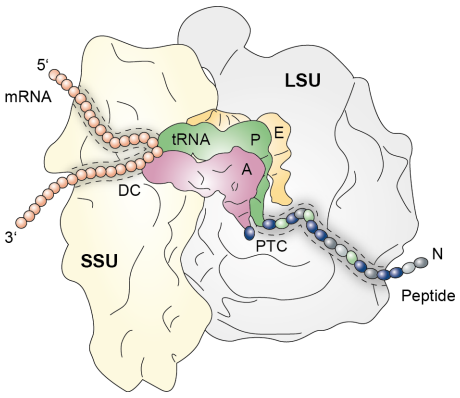


Figure 1 | The ribosome. Schematic representation of the ribosome with SSU and LSU displayed in yellow and grey, respectively. The mRNA path and polypeptide exit tunnel are indicated with dotted lines. A-, P- and E-tRNAs are colored in purple, green and orange, respectively. Figure modified from (Lafontaine and Tollervey, 2001).

Ribosomes utilize tRNAs with the help of translation factors to transfer information from the DC to the PTC. The binding sites for tRNAs are formed by the interface sides of both subunits and are named aminoacyl-(A), peptidyl-(P) and exit-(E) site, according to the state of tRNA they are housing (Figure 1). During the translation process tRNAs move sequentially from A- through P- to E-site. The A-site binds the aminoacyl-tRNA (A-tRNA) carrying the next amino acid to be incorporated into the growing peptide chain that is bound to the peptidyl-tRNA (P-tRNA) located

in the P-site. Deacylated tRNAs occupy the E-site before dissociating from the ribosome. Polypeptides are elongated from the N- to the C-terminus at the PTC, which lies in the center of the LSU. To reach their destination in the cell all proteins need to pass through a tunnel that emanates adjacent to the PTC and spans the LSU until it emerges in the cytosol (Figure 1). Decoding and peptidyl transfer are the two fundamental processes on the ribosome and hence require a more detailed description.

Decoding. During decoding the ribosome selects cognate tRNAs at the A-site while rejecting near-cognate tRNAs. The process relies on base pairing of the tRNA anticodon with the mRNA codon and is the single step in translation that links the genetic code to amino acid selection (reviewed in Ogle and Ramakrishnan, 2005; Rodnina and Wintermeyer, 2001; Zaher and Green, 2009). Codon recognition results in conformational changes in the universally conserved SSU rRNA residues A1492, A1493 and G530 (*Escherichia coli* numbering). This allows them to interact with the first, second but not third position of the mRNA-tRNA minihelix minor groove (Ogle et al., 2001) in the form of so called A-minor motifs (Nissen et al., 2001). These contacts monitor the correct Watson-Crick geometry of the codon-anticodon base pair in the first two positions but allow wobble pairs (G·U) at the third position and trigger a large scale domain closure of the SSU to allow the subsequent steps of the translation cycle (Ogle et al., 2002). In addition, parts of the tRNA body distant from the anticodon also contribute to accuracy during decoding by influencing the energetics of tRNA distortion during A-site binding in the context of translation factors (Schmeing et al., 2011; 2009). The model of decoding has recently been challenged by the observation that near-cognate tRNAs can induce a closed conformation of the ribosome similar to cognate tRNAs. In this scenario the ribosome forces G·U pairs of near-cognate tRNAs into an unfavorable Watson-Crick geometry and the associated energetically penalty is thought to allow tRNA discrimination and ensures fidelity during decoding (Demeshkina et al., 2012).

Peptidyl transfer. The reaction proceeds via nucleophilic attack of the A-tRNA α-amine on the carbonyl carbon of the P-tRNA ester (Leung et al., 2011). The PTC interacts with the 3'-CCA ends of A- and P-tRNAs and positions the substrates for attack, while preventing P-tRNA hydrolysis by the omnipresent water molecules when the A-site is empty. This occurs via substrate induced fit of the LSU rRNA (Schmeing et al., 2005; Voorhees et al., 2009). Proper orientation of the reaction partners contributes largely to the catalytic power of the ribosome ($2x10^7$-fold enhancement compared to the un-catalyzed reaction), which is thought to function as an entropy trap (Sievers et al., 2004). However, the transition state and thus the mechanism of ribosome-catalyzed peptide transfer differs significantly from the un-catalyzed reaction (Kingery et al., 2008), which argues against a purely entropic effect. The exact nature of the transition state(s) (Hiller et al., 2011; Kuhlenkoetter et al., 2011) and the precise role of the 2'OH of P-tRNA in the reaction on the ribosome are still hot topics in the field.

Both DC and PTC are built up from rRNA (Nissen et al., 2000; Ogle et al., 2001). In fact, RNA plays the starring role in the two basic activities of the ribosome, making it a ribozyme (Cech, 2000). This indicates that the modern ribosome is the living fossil of a primitive RNA catalyst originating from an ancient RNA world (Fox, 2010; Noller, 2012).

1.2 Ribosomes in the Three Domains of Life

Ribosomes are universally conserved in their function throughout the three domains of life (Bacteria, Archaea and Eukarya). Despite this conservation their composition and size differ significantly (Melnikov et al., 2012). While bacterial and archaeal 70S ribosomes are composed of

a small 30S and a large 50S subunit (Figure 2a,b), eukaryotic small 40S and large 60S subunits are considerably larger and together form the 80S ribosome (Figure 2c). The size difference is due to additional rRNA in the form of expansion segments (ES) as well as many eukaryote-specific r-proteins and r-protein extensions (Klinge et al., 2012; Melnikov et al., 2012; Wilson and Cate, 2012). This increased structural complexity is thought to reflect the more complex processes of ribosome biosynthesis, translation initiation and regulation operating in the eukaryotic domain of life. Beside the specific parts, all ribosomes possess a conserved core that contains all structural features (DC, PTC, polypeptide exit tunnel, tRNA and translation factor binding site) necessary for the basic functions during translation (Figure 1). The core is roughly formed by 4400 nucleotides (nts) of rRNA and 34 r-proteins (Melnikov et al., 2012).

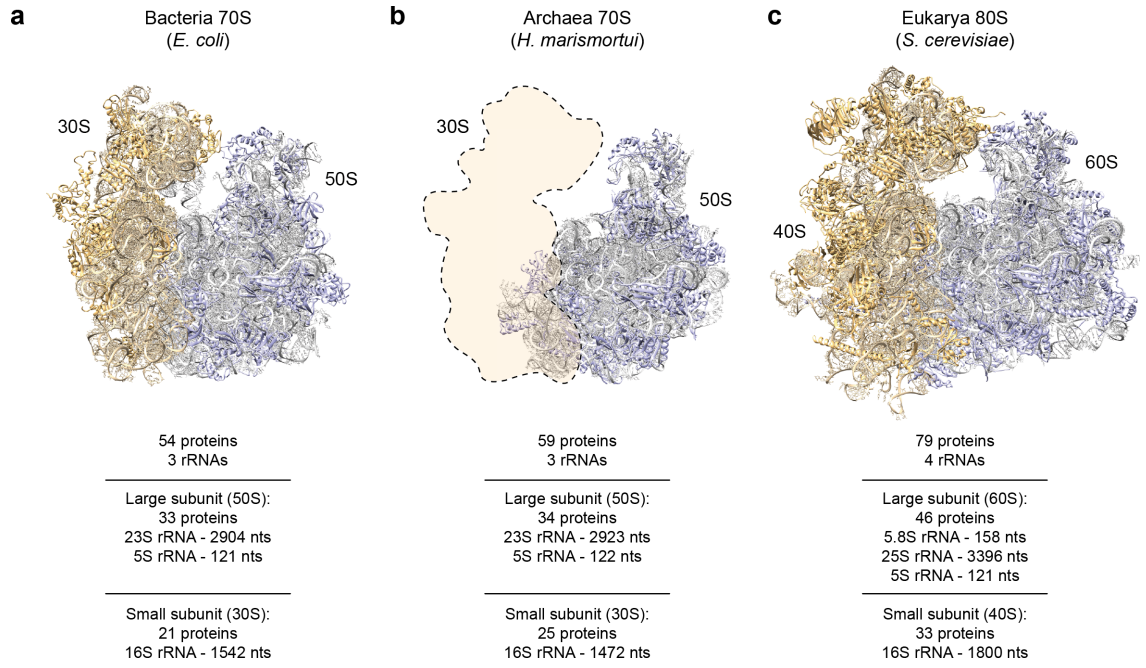


Figure 2 | Composition of bacterial, archaeal and eukaryotic ribosomes. (**a**) Structure of the bacterial 70S ribosome from *E. coli* (Dunkle et al., 2011) with rRNA/r-proteins colored in orange/light tan and violet/grey for the small and large ribosomal subunit, respectively. (**b**) Structure of the archaeal large ribosomal subunit from *Haloarcula marismortui* (Kavran and Steitz, 2007) with rRNA/r-proteins colored as in (a). Position of the small ribosomal subunit is indicated schematically. (**c**) Structure of the eukaryotic 80S ribosome from *Saccharomyces cerevisiae* (Ben-Shem et al., 2011) with rRNA/r-proteins colored as in (a).

Differences in size are also present within the eukaryotic domain with higher eukaryotes tending to have larger ribosomes. While all eukaryotes contain the same set of about 80 core r-proteins, this divergence is largely achieved via variations in rRNA length. A striking example is given by comparing the lower and higher eukaryotic ribosomes from *S. cerevisiae* and *Homo sapiens*, respectively. The human ribosome has a molecular mass of 4.3 MDa and contains ~7200 nts of rRNA, compared to the 3.3 MDa and ~5500 nts in yeast. Other metazoan species like the fruit fly *Drosophila melanogaster* (~6300 nts of rRNA) take an intermediate position on this list. Interesting possibilities are that the extended rRNA structures are related to the phenomenon of localized translation in the nervous system (Wang et al., 2010) or translational control during metazoan development (Richter and Lasko, 2011).

1.3 Ribosome Structures

Current understanding of the ribosome architecture and function is based on structural studies to a large extent. Detailed insights came from crystal structures of the bacterial SSU from *Thermus thermophilus* (Schluenzen et al., 2000; Wimberly et al., 2000), LSUs from *H. marismortui* (Ban et al., 2000) and *Deinococcus radiodurans* (Harms et al., 2001), as well as complete 70S ribosome structures from *E. coli* and *T. thermophilus* (Schuwirth et al., 2005; Selmer et al., 2006; Yusupov et al., 2001). These studies revealed the complex architecture of the ribosome resulting from the interactions of r-proteins and rRNA and constitute the basis for X-ray structures of the bacterial 70S in complex with elongation and release factors (Gao et al., 2009; Korostelev et al., 2008; Laurberg et al., 2008; Petry et al., 2005; Pulk and Cate, 2013; Schmeing et al., 2009; Tourigny et al., 2013; Weixlbaumer et al., 2008; Zhou et al., 2013). Bacterial and archaeal crystal structures have provided unparalleled insights into the translation mechanism (Schmeing and Ramakrishnan, 2009; Voorhees and Ramakrishnan, 2013) as well as the inhibition of ribosomes by antibiotics (Wilson, 2009). Moreover, the knowledge about RNA folding and especially structural RNA motifs (e.g. A-minor or kink-turns (KT)) grew substantially with the first crystal structures of the ribosome (Klein et al., 2001; Nissen et al., 2001; Noller, 2005). More recently, crystallography succeeded in solving structures of lower eukaryotic ribosomes, namely the 40S and 60S subunits from *Tetrahymena thermophila* (Klinge et al., 2011; Rabl et al., 2011) and the entire 80S yeast ribosome from *S. cerevisiae* (Ben-Shem et al., 2010; 2011). These works provided insights into the architecture of eukaryote-specific elements and their interaction with the ribosomal core at atomic detail.

The high-resolution structures obtained from X-ray crystallography are complemented by cryo-electron microscopy (cryo-EM) reconstructions, which have proven to be particularly useful for the visualization of ligands bound to the ribosome (Becker et al., 2009; 2011; 2012; Beckmann et al., 2001; Halic et al., 2004; 2006). The technique offers several advantages over crystallography. It requires only small amounts of (non-crystalline) sample and is especially suited for the investigation of large macromolecular complexes in a nearly native environment. Furthermore, computational procedures during data processing can correct for conformational or sample specific heterogeneity. In fact, the interactions of elongation factors with the bacterial ribosome have initially been visualized by means of cryo-EM (Agrawal et al., 1998; Stark et al., 1997). Functional interpretation of the complexes became possible with better resolution but did not reach atomic detail (Connell et al., 2007; Schuette et al., 2009) and accordingly, crystallography was needed to reveal molecular mechanisms (Gao et al., 2009; Schmeing et al., 2009; Voorhees et al., 2010). Likewise, the first visualization of a eukaryotic ribosome by cryo-EM dates back to 1996 (Verschoor et al., 1996). Since then several cryo-EM studies aimed at mapping (and modeling) eukaryote-specific parts in ribosome reconstructions from different species (Chandramouli et al., 2008; Hashem et al., 2013a; Spahn et al., 2001; Taylor et al., 2009).

Constant improvement of technical equipment, processing techniques and computational power resulted in cryo-EM reconstructions with ever improving resolution. A concrete example is given by the available cryo-EM reconstructions of the *S. cerevisiae* ribosome that range from 15.4 Å in 2001 (Beckmann et al., 2001; Spahn et al., 2001) over 11.7 Å in 2004 (Spahn et al., 2004a), 9.9 Å in 2006 (Andersen et al., 2006) and 6.1 Å in 2009 (Becker et al., 2009) to 4.5 Å in 2013 (Bai et al., 2013). In contrast to this, structural information for the more complex translational apparatus of higher eukaryotes, like mammalians is still limited. The medium to low resolution (9 to 29 Å) of the available cryo-EM reconstructions has so far prohibited the generation of complete molecular models (Budkevich et al., 2011; Chandramouli et al., 2008; Dube et al., 1998a; 1998b; Ménétret et al., 2000; Morgan et al., 2000; Spahn et al., 2004b). Cryo-EM seems to be the best choice to

obtain high-resolution structures of mammalian ribosomes since the increased complexity of these particles very likely makes them challenging targets for crystallography. Accordingly, initial crystal structures of the mammalian 40S subunit in complex with initiation factors only reached resolutions of 7.9 to 9 Å (Lomakin and Steitz, 2013). Currently cryo-EM and single particle reconstructions can reach near atomic resolution even for an asymmetric assembly like the ribosome and thus start to rival crystallographic studies. Moreover, recent developments like direct electron detectors and beam induced motion correction (Li et al., 2013) are very likely to push the resolution for ribosomes below 4 Å soon. This would allow dissection of molecular mechanisms operating on the ribosome at atomic detail, combined with the advantages offered by cryo-EM.

1.4 Ribosomal RNA

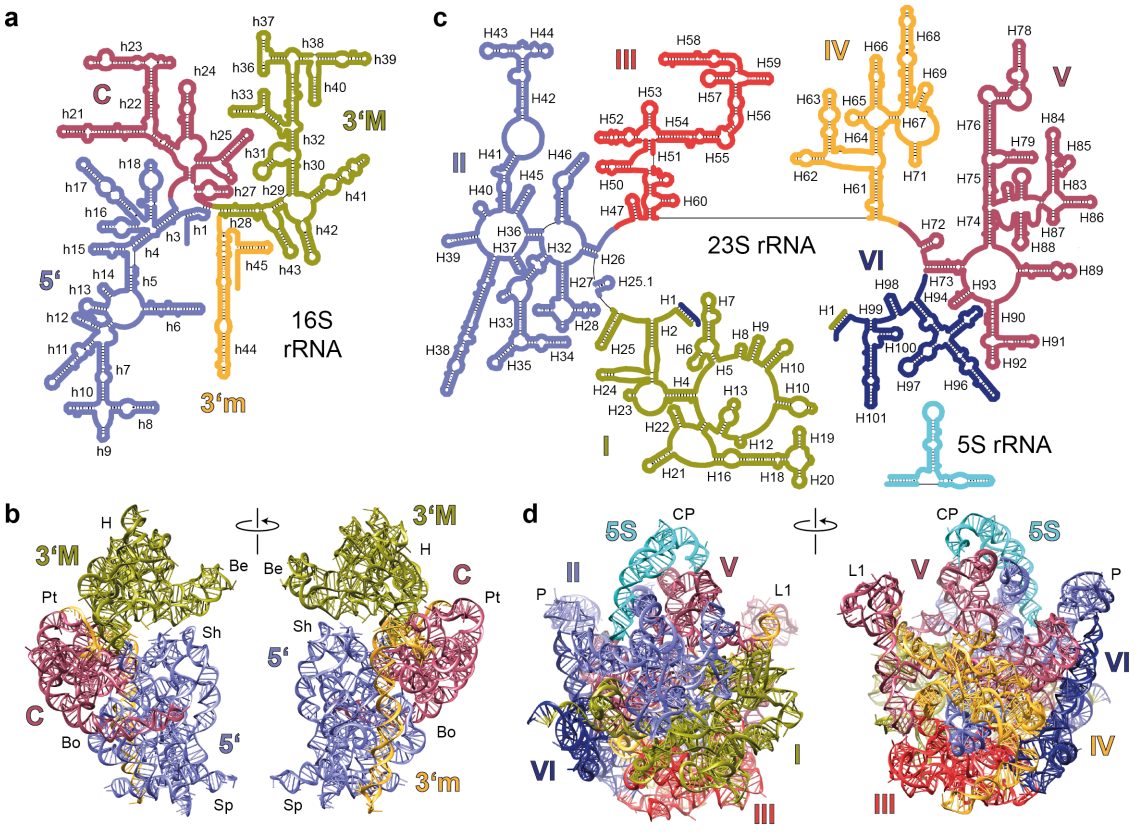


Figure 3 | Ribosomal RNA. (**a**,**b**) Secondary structure diagram (**a**) and three-dimensional fold (**b**) of the 16S rRNA from *E. coli* (Dunkle et al., 2011). Domains are colored distinctly and rRNA helices are numbered. (**c**,**d**) Secondary structure diagram (**c**) and structure (**d**) of the 23S/5S rRNAs from *E. coli* (Dunkle et al., 2011). Domains are colored individually and helices are numbered. Be, beak; Bo, body; CP, central protuberance; H, head; L1, L1-stalk; P, L7/L12(P)-stalk; Pt, platform; Sh, shoulder; Sp, spur.

Bacterial 70S ribosomes are composed of three rRNAs (SSU: 16S, LSU: 23S and 5S). These are enlarged in eukaryotes and due to a cleavage site in the terminal loop of LSU helix 10 (H10) an additional rRNA piece (5.8S) is found, collectively resulting in the four rRNAs of 80S ribosomes (SSU: 18S, LSU: 28S, 5.8S and 5S). In several eukaryotic species the large subunit rRNAs are further subdivided in smaller pieces. Examples from *Drosophila* include cleavage of 5.8S rRNA in the tip of

H9 to produce the 2S rRNA and the generation of 28Sα and 28Sβ (Tautz et al., 1988; Ware et al., 1985). Functional significance of these additional processing steps is unknown at present.

The SSU rRNA can be subdivided into four domains (5' and 3' minor, 3' major and central), which are clearly discernable and constitute the conserved structural landmarks (head, body and platform) of the subunit (Figure 3a,b). The structural flexibility of the individual domains with respect to each other is a prerequisite for the rotation and head swivel movements of the SSU during translation (Frank and Agrawal, 2000; Schuwirth et al., 2005). In contrast to this, the seven rRNA domains (I to VII) of the large ribosomal subunit (5S rRNA as domain VII) are intricately interwoven with each other, resulting in a single compact three-dimensional entity (Figure 3c,d). Structural landmarks of the LSU include the central protuberance (CP) as well as the flexible L1 and L7/L12(P)-stalks.

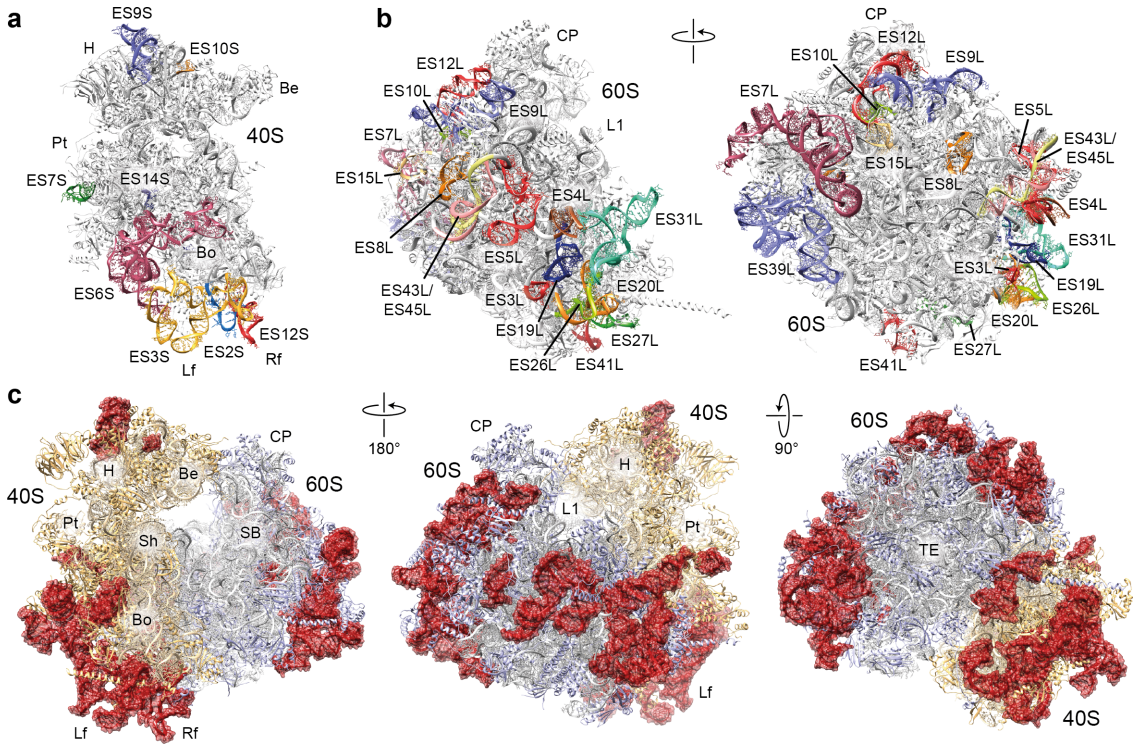


Figure 4 | rRNA Expansion segments. (**a**,**b**) Structures of the yeast 40S (**a**) and 60S (**b**) subunits (Ben-Shem et al., 2011) with ES colored individually. Due to flexibility, helix ES7L-A and the majority of ES27L are not contained in the structure. (**c**) Structure of the yeast 80S ribosome (Ben-Shem et al., 2011) viewed from the A-site (left), from the E-site (middle) and from the tunnel exit (TE) side (right). ES are highlighted in red, remaining rRNA and r-proteins are colored in light tan/orange and grey/violet for the 40S and 60S subunit, respectively. An extended rRNA ES definition, based on (Gerbi, 1996) is used throughout this thesis (for details see Anger et al., 2013). Lf, left foot; Rf, right foot; SB, P-stalk base; TE, tunnel exit.

Expansion segments (ES). Eukaryotic ribosomes are significantly larger than their bacterial counterparts and this size difference is mainly due to additional rRNA portions called expansion segments (ES) (Figure 4). These structures are also the main reason for the increased mass of higher eukaryotic ribosomes in comparison to representatives from lower eukaryotic species. ES are a subset of variable regions (VR) within rRNA that are not evolutionary conserved. They disrupt the common rRNA core at the same positions but vary in size between different organisms, suggesting a common evolutionary origin (Cannone et al., 2002). In principle all eukaryotic ribosomes share the same general ES topology and species differences arise mainly by

variations in size. However exotic examples regarding ES topology from *Mycobacteria* (Shasmal and Sengupta, 2012) or *Trypanosoma* are known (Gao et al., 2005; Hashem et al., 2013a). On the SSU the ES cluster mainly at the lower part of the structure, which results in a remodeled foot region (Figure 4a) (Ben-Shem et al., 2011; Rabl et al., 2011). The tight interaction of ES3S and ES6S via base pairing creates the so-called left foot, while the right foot of the eukaryotic SSU is formed by ES12S. The majority of the remaining ES are also found in the lower part of the particle, with the exception of ES9S and ES10S, which are located in the head domain (Figure 4a). On the LSU ES are mainly found in two clusters on the back and side of the particle. The first one is positioned behind the P-stalk. It is organized around ES7L and ES39L and also contains ES9L, ES10L, ES12L and ES15L. The second one, located behind the L1-stalk, is formed by ES31L with the surrounding ES3L, ES4L, ES5L, ES19L, ES20L and ES26L (Figure 4b) (Ben-Shem et al., 2011; Klinge et al., 2011). In general, ES are found on the surface of the ribosome, leaving the intersubunit sides and functional sites like the translation factor binding site, or the tunnel exit rather unaffected (Figure 4c). ES are thought to originate from slippage events during replication of the rRNA gene array that eventually lead to the enlargement of the rRNA. This idea is based on the finding, that many ES are found to be “cryptic simple”, meaning that they contain stretches with degenerated, repetitive motifs (Hancock et al., 1988). Cryptic simple sequences are absent from the conserved rRNA core sequences since slippage events during replication are likely incompatible with retaining the ribosomal core structure and hence functionality in most cases. Interestingly, the 18S rRNA of *D. melanogaster* does not contain cryptically simple stretches to the same extent as the LSU rRNA (Tautz et al., 1988). This could be seen as an indication that ESs of the SSU are subject to constraints due to some important functions (e.g. during initiation of translation). Little is known about the function of ES and in fact it is even possible that the majority of ES don’t have a functional role, but are only tolerated within the rRNA structure because they are not interfering with ribosome function (Clark, 1987). Consistent with this idea is the finding that bacterial ribosomes are tolerant towards rRNA insertions at various positions and that several of these coincide with the location of ES in the rRNA structure (Yokoyama and Suzuki, 2008). Generally, ES could fulfill a function on the DNA or RNA level. On the DNA level they might serve as hotspots for homologous recombination, which is a prerequisite to maintain a homogeneous rRNA gene population. Ribosomal DNA (rDNA) clusters are present in multiple copies in the cell and despite this multiplicity the encoded rRNA molecules are homogeneous. The rDNA genes are thought to undergo continual rounds of unequal crossover to maintain this homogeneity (Eickbush and Eickbush, 2007). On the RNA level eukaryote-specific functions could arise from the ES itself of from proteins that specifically bind to them. More concrete indications of possible function are available for ES7L and ES27L, the two largest ES of the LSU. ES7L deletions are lethal in *S. cerevisiae* (Jeeninga et al., 1997) and similar results have been obtained for ES27L in *T. thermophila* and *S. cerevisiae* (Jeeninga et al., 1997; Sweeney et al., 1994). Moreover, in *Tetrahymena* the lethal phenotype can be rescued by insertion of ES27L sequences from other species but not unrelated rRNA stretches (Sweeney et al., 1994). The observation that ES27L is flexible and can adopt different conformations in cryo-EM reconstructions of the *S. cerevisiae* ribosome led to the initial suggestion that it might dynamically control access of ligands to the polypeptide tunnel exit (Beckmann et al., 2001). In the meantime ES27L has been observed to interact with a series of important factors, such as the ribosome-associated complex (RAC) (Leidig et al., 2013), the biogenesis factor Arx1 (Bradatsch et al., 2012; Greber et al., 2012b) and the membrane protein ERj1 (Blau et al., 2005). Several ES (including ES7L and ES27L) have also been shown to be selectively cleaved during apoptosis (Houge et al., 1993; 1995) and are important for correct rRNA processing during ribosome biogenesis (Jeeninga et al., 1997). To summarize, the list of data pointing towards a functional role of some ES is growing and the two largest ES of the LSU (ES7L and ES27L) are emerging prime candidates.

1.5 Ribosomal Proteins

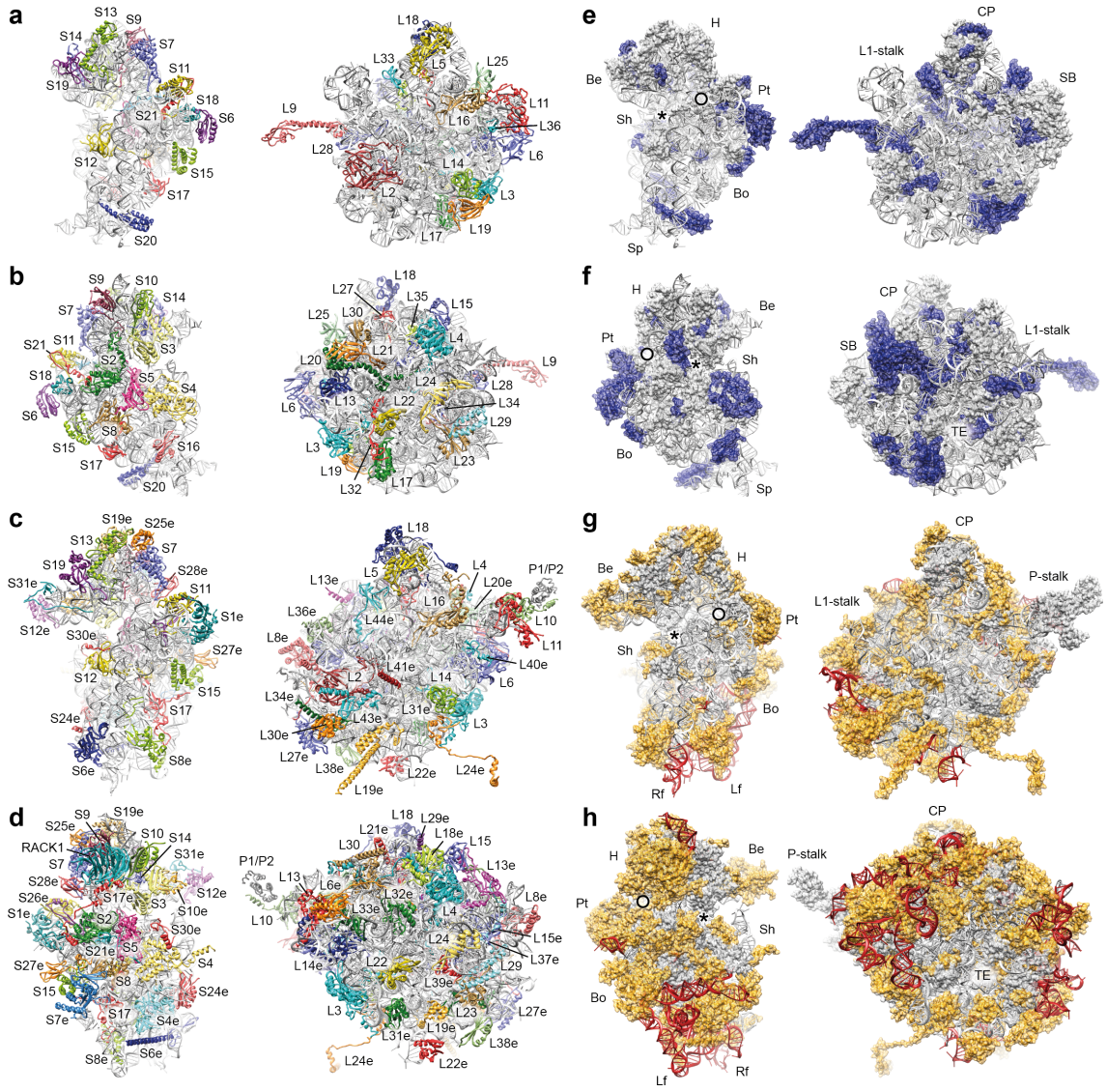


Figure 5 | Protein architecture of bacterial and eukaryotic ribosomes. (**a**,**b**) Interface (**a**) and solvent (**b**) view of the bacterial 30S (left) and 50S (right) ribosomal subunits from *E. coli* (Dunkle et al., 2011) with rRNA shown in grey and r-proteins colored individually. (**c**,**d**) Interface (**c**) and solvent (**d**) view of the eukaryotic 40S (left) and 60S (right) ribosomal subunits from *S. cerevisiae* (Ben-Shem et al., 2011). rRNA and r-proteins are colored as in (a,b). (**e**,**f**) Interface (**e**) and solvent (**f**) view of the bacterial subunits as in (a,b) with bacterial-specific r-protein elements highlighted in blue. (**g**,**h**) Interface (**g**) and solvent (**h**) view of the yeast subunits as in (c,d) with eukaryote-specific r-protein elements colored in orange. ES are highlighted in red. The recently revised nomenclature for r-proteins is used throughout this thesis (Ben-Shem et al., 2011; Jenner et al., 2012). In contrast to the original proposal, eukaryotic r-protein P0 is named L10, as suggested by Liljas, Moore and Yusupov (www.elsevierblogs.com/currentcomments/?p=686). Due to flexibility r-proteins L1 (*E. coli* and *S. cerevisiae*), L31 (*E. coli*) and stalk proteins L10, L7/L12 (*E. coli*) are not contained in the structures. The same holds true for the weakly bound bacterial r-protein S1, which is located at the mRNA exit site on the 30S subunit (Sengupta et al., 2001). mRNA entry and exit sites on the 30S and 40S subunits are indicated with an asterisk and circle, respectively.

In bacteria, such as *E. coli*, ribosomes contain ~54 r-proteins (SSU: 21; LSU: 33) (Figure 5a,b), including 34 r-proteins that are universally conserved across all kingdoms of life. In comparison, ~80 core r-proteins (SSU: 33; LSU: 47) are found in eukaryotic ribosomes. Of these, 34 (SSU: 13; LSU: 21) are shared with archaea, which in total contain 55-69 (SSU: 24-28; LSU: 31-41) r-proteins, depending on the individual species. This leaves 12 (SSU: S7e, S10e, S12e, S21e, RACK1; LSU: L6e, L20e, L22e, L27e, L28e, L29e, L36e) r-proteins that are exclusive for eukaryotic ribosomes (Supplementary Tables S1 and S2) (Figure 5c,d). It has to be noted that the above numbers only give a general idea about the r-protein numbers in the three domains of life, since individual species might have lost some r-protein or contain additional ones. For instance an extra r-protein, Thx is found in the thermophilic bacterium *T. thermophilus* (Choli et al., 1993), while r-protein L28e is missing in *S. cerevisiae* due to loss of its gene in this organism (Lecompte et al., 2002). Both of these r-proteins serve as glues for rRNA structures. Thx is buried in 16S rRNA to stabilize the structure (Wimberly et al., 2000) and L28e used to anchor a helical part of ES7L (the ES7L-A helix) to the body of the ribosome (Klinge et al., 2011). In the absence of L28e, ES7L-A is highly flexible and could not be resolved in the yeast 80S X-ray structure (Ben-Shem et al., 2011).

Bacterial r-proteins are mainly bound to the surface of the particle with globular domains that are connected to flexible linkers, which in turn are able to weave through internal parts of the ribosome and closely interact with rRNA (Brodersen et al., 2002; Klein et al., 2004; Wilson and Nierhaus, 2005). However, the linkers are not primarily used to interact with neighboring proteins (Figure 5e,f). In contrast to this, eukaryote-specific r-protein elements (specific r-proteins and extensions) are extensively used to establish tertiary contacts with ES as well as other r-proteins. For instance L6e, L27e, L29e have architectural roles in stabilizing contacts between ES7L-ES39L, ES31L-ES20L/ES26L and ES9L-ES12L, respectively and inter-protein secondary structure elements such as β-sheets are frequently used to stabilize the eukaryotic ribosome (Ben-Shem et al., 2011; Klinge et al., 2011). The intertwined nature of the eukaryote-specific elements reveals co-evolution of rRNA and r-proteins as a general feature of 80S ribosomes. Examples include r-protein S7e and the base of ES6S (Ben-Shem et al., 2011; Rabl et al., 2011), the shortened h33 at the SSU beak and its replacement by S10e/S12e (Ben-Shem et al., 2011; Rabl et al., 2011), as well as the large concentration of eukaryote-specific elements on the back of the LSU, where ES7L, ES39L are intricately interwoven with L6e, L14e, L28e, L32e, L33e plus extensions of L4, L13 and L30 (Ben-Shem et al., 2011; Klinge et al., 2011). The eukaryote-specific additions and inter-connections are mainly located on the surface of the ribosome, with the intersubunit sides and tunnel exit being largely conserved between bacteria and eukaryotes (Figure 5g,h). Several eukaryote-specific proteins are involved in the formation of intersubunit bridges and these often involve eukaryote-specific rRNA elements on the other subunit. Examples are eukaryote-specific bridge (eB) 8 (ES31L - S1e), eB11 (ES41L - S8e), eB12 (L19e - ES6S) and eB13 (L24e - S6e/h10) (Ben-Shem et al., 2011), that together with eB9 (L30e - S13e) (Halic et al., 2005), eB10 (H63 - h11/S8e) (Spahn et al., 2001) and eB14 (L41e - h27/h44/h45) (Ben-Shem et al., 2011) result in a nearly doubled interaction surface between the subunits in eukaryotes in comparison to bacterial 70S ribosomes. An interesting feature of eB12 and eB13 is that the involved proteins (L19e and L24e, respectively) bind to the solvent exposed side of the 40S with extended helical parts which is reminiscent of the bridge formed by L31 between the LSU central protuberance and the SSU head of the bacterial ribosome (Jenner et al., 2010). Most of the eukaryote-specific bridges are located at the periphery of the subunit interface, with eB14 (formed by L41e) being a striking exception. L41e is the smallest r-protein and with 25 amino acids (aa) even the smallest protein in the yeast genome. The protein is completely surrounded by rRNA at the center of the subunit interface (Figure 5c). It was noticed to be present in the yeast 80S ribosome (Ben-Shem et al., 2011) but absent in structures of the individual subunits from *Tetrahymena* (Klinge

et al., 2011; Rabl et al., 2011). Interestingly, the L41e binding pocket, formed by rRNA of both subunits is conserved in bacteria and archaea, but no corresponding protein has been identified to date.

Although a wealth of information is available on the architecture of eukaryotic r-proteins and their interaction within the ribosome, only a few specific functions are known. RACK1 on the SSU functions in cell signaling by serving as a binding platform for protein kinase C (PKC) (Grosso et al., 2008b; Sharma et al., 2012) and several other factors such as Scp160 (Coyle et al., 2009; Nilsson et al., 2004). PKC activity on the ribosome has been implicated in ribosome assembly (Ceci et al., 2003) and translational regulation (Grosso et al., 2008b; 2008a). However, detailed insights on the underlying mechanisms are currently lacking. S6e is another r-protein involved in translational control. It is phosphorylated by S6-kinases upon activation of the mTOR signaling pathway, which regulates cell growth and division (Meyuhas and Dreazen, 2009; Ruvinsky and Meyuhas, 2006). Two r-proteins, S31e and L40e, are synthesized as fusions to ubiquitin (Finley et al., 1989). Interestingly, these fusion proteins are located near the decoding site (S31e) and the translation factor binding site (L40e) of the ribosome and uncleaved ubiquitin moieties would obstruct essential functions during translation (Ben-Shem et al., 2011; Klinge et al., 2011; Rabl et al., 2011). At present it is not clear if S31e and L40e are incorporated with ubiquitin during ribosome biogenesis. S31e, together with S30e, has a function during the initiation phase of translation by binding of eukaryotic initiation factor (eIF) 1A to the 40S subunit (Weisser et al., 2013). Likewise, eukaryote-specific r-proteins S1e, S26e and S27e form a docking site for eIF3 (Hashem et al., 2013b). Other specific functions are known for r-proteins S19e and S25e, which are part of binding site for the yeast-specific elongation factor eEF3 (Andersen et al., 2006) and L38e and L40e have been implicated to regulate the specific translation of a subset of homeobox and viral mRNAs, respectively (Kondrashov et al., 2011; Lee et al., 2013).

1.6 Ribosome Biogenesis

Bacterial ribosome biogenesis is driven by self-assembly and assisted by many nonribosomal factors that render the process more efficient in vivo (reviewed in Connolly and Culver, 2009; Shajani et al., 2011). The self-assembly capacity of bacterial ribosomes is underscored by the fact that biogenesis factors are not strictly required and functional ribosomes can be reconstituted *in vitro* (Nomura, 1973). The latter remains challenging for eukaryotic ribosomes and the majority of the numerous assembly factors are essential in yeast (Dinman, 2009). An additional layer of complexity is given by the compartmentalization of the process, which occurs in the nucleolus, the nucleoplasm and the cytoplasm of eukaryotic cells (Figure 6) (reviewed in Kressler et al., 2010; Panse and Johnson, 2010).

Eukaryotic ribosome biogenesis starts in the nucleolus with the transcription of a large rRNA precursor containing rRNA pieces from both subunits (18S, 28S, 5.8S) by RNA polymerase I. 5S rRNA is transcribed separately by RNA polymerase III in the nucleoplasm from where it delivered to the nucleolus together with r-proteins L5 and L18 (L11 and L5, respectively, according to the old yeast nomenclature). These and all other r-proteins are translated in the cytoplasm (from mRNA transcribed by RNA polymerase II) and need to be imported into the nucleus to become available in the biogenesis pathway. Assembly of the ribosomal subunits involves a complicated series of rRNA processing and modification steps, that together with the incorporation of r-proteins ultimately lead to the formation of export-competent pre-40S/60S particles that independently leave the nucleus through the nuclear pore complexes (Figure 6). In yeast, about 75 small nucleolar ribonucleoprotein particles (snoRNPs) responsible for rRNA modifications

(methylation, pseudouridylation) and more than 200 non-ribosomal factors are involved in the biogenesis process (Fromont-Racine et al., 2003; Henras et al., 2008). The latter include many GTPases, AAA family ATPases, RNA helicases, chaperones and kinases, which suggests that a significant amount of remodeling is required to assemble functional ribosomal subunits in eukaryotes. This is very likely related to the intertwined layer of eukaryote-specific elements in the 80S ribosome (Ben-Shem et al., 2011). The current understanding of eukaryotic ribosome biogenesis is largely based on experiments performed with *S. cerevisiae* and knowledge about the processes in higher eukaryotes such as human lags far behind. Given the increased complexity of higher eukaryotic ribosomes, biogenesis probably includes special features that cannot be found in yeast. For instance, precursor rRNA processing in humans involves numerous factors that have no yeast homolog (Tafforeau et al., 2013) and a nuclear export route specific for the LSU in higher eukaryotic cells has been identified (Wild et al., 2010). Other features that might influence the complexity of ribosome biogenesis in higher eukaryotes are a unique link to stress response (Zhang and Lu, 2009), as well as differences in the rDNA repeat organization (Prokopowich et al., 2003; Richard et al., 2008).

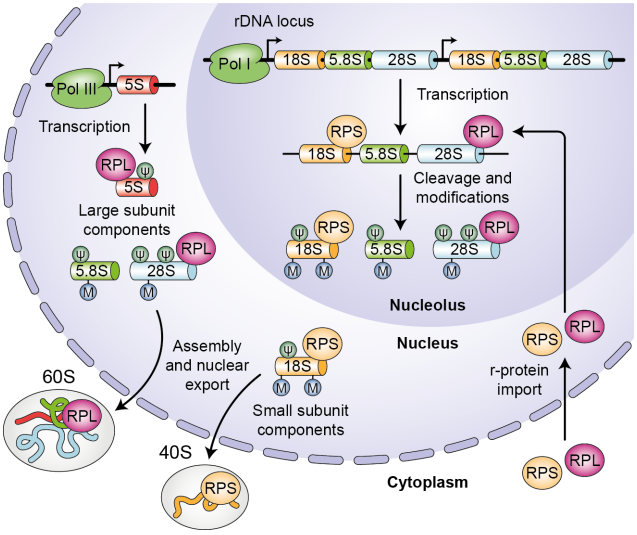


Figure 6 | Ribosome biogenesis. M, methylation; ψ, pseudouridylation. Pol, RNA polymerase, RPL, r-protein of the LSU; RPS, r-protein of the SSU. Figure modified from (Xue and Barna, 2012).

Structural information regarding ribosome biogenesis in eukaryotes is still very limited and all currently available structures are exclusively from lower eukaryotic species. The majority of the studies were performed with mature 60S subunits and artificially re-bound biogenesis factors (Gartmann et al., 2010; Greber et al., 2012b; Klinge et al., 2011; Sengupta et al., 2010) and cryo-EM structures of endogenous late pre-40S/60S particles have only become available recently (Bradatsch et al., 2012; Strunk et al., 2011). They illustrate how assembly factors mask functional sites of the premature subunits and reveal a connection between ES27L and the biogenesis factor Arx1 (Bradatsch et al., 2012; Greber et al., 2012b).

1.7 Translation Mechanism

Translation proceeds through four stages: initiation, elongation, termination and recycling (Rodnina and Wintermeyer, 2009; Schmeing and Ramakrishnan, 2009). Elongation lies at the heart of the translation process and is highly conserved in contrast to the remaining phases, which differ significantly between bacteria and eukaryotes (Figure 7). Although being prokaryotes, archaea have more complex initiation, termination and recycling pathways than bacteria that more resemble the situation in eukaryotes (Benelli and Londei, 2011; Franckenberg et al., 2012).

Initiation. Initiation includes all steps required to form an elongation-competent ribosome with initiator-tRNA bound to the P-site on the mRNA start codon. In bacteria it involves three initiation factors (IF1, IF2, IF3) and the Shine-Dalgarno (SD) sequence (Shine and Dalgarno, 1974) within the mRNA to be translated (reviewed in Simonetti et al., 2009). The process is driven in large parts by base pairing of the SD sequence with the 5' end of 16S rRNA (the SD interaction) (Kaminishi et al., 2007), which directs positioning of the 30S subunit over the start codon.

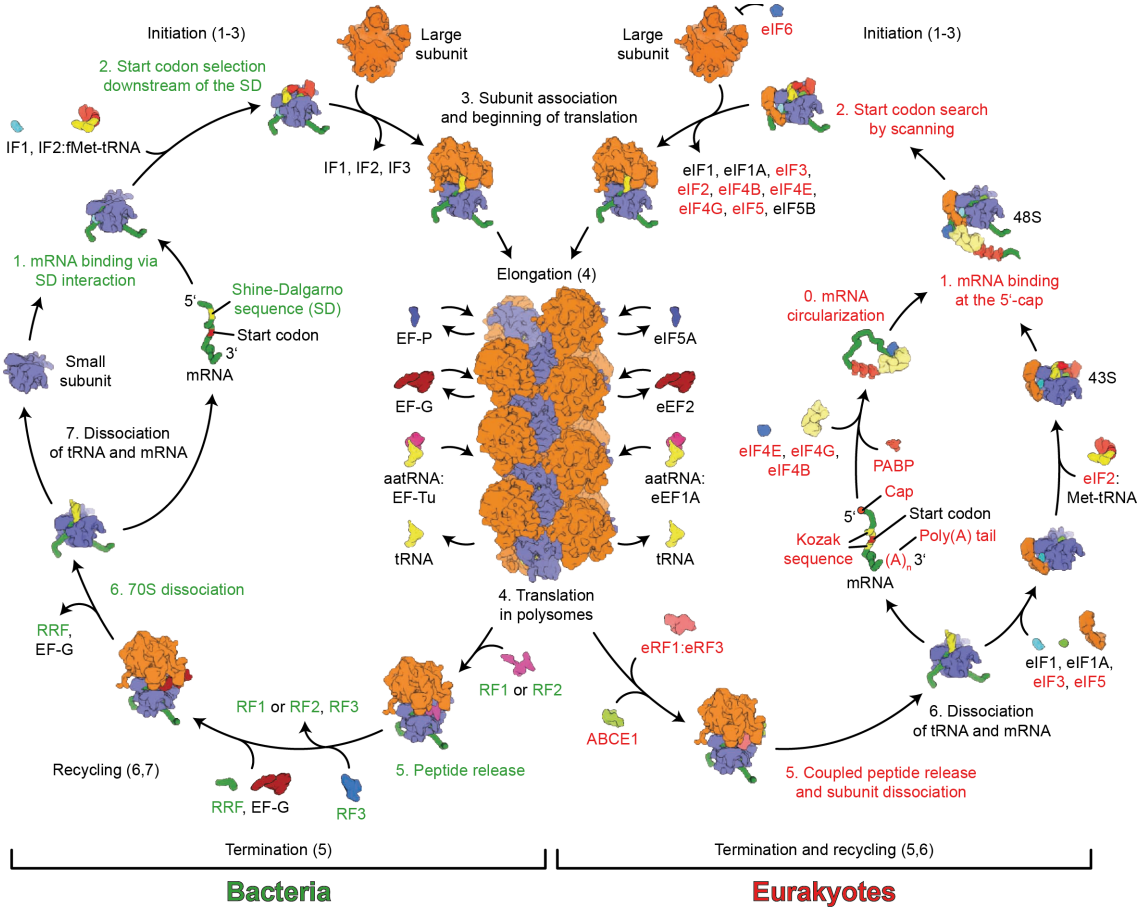


Figure 7 | Translation cycle in bacteria and eukaryotes. Conserved factors are labeled in black, while bacterial and eukaryote-specific factors are indicated in green and red, respectively. Figure modified from (Melnikov et al., 2012).

In contrast to this, initiation in eukaryotes is far more complicated and involves 13 core eukaryotic initiation factors (eIFs), some of them being large multisubunit complexes (reviewed in Hinnebusch and Lorsch, 2012; Jackson et al., 2010; Voigts-Hoffmann et al., 2012). The 40S subunit is initially bound by eIF1 (in the P-site), eIF1A (in the A-site), eIF3 and probably eIF5. Within this complex eIF1 recruit the ternary complex (TC), consisting of eIF2, initiator-tRNA and GTP to form a 43S pre-initiation complex (PIC). The 43S PIC binds an mRNA via interaction with eIF4 and the resulting complex starts 5' to 3' scanning along the mRNA until a start codon is encountered. GTP hydrolysis by eEF2 is the key step in initiation and is controlled by eIF5 (stimulates the reaction) and eIF1 (blocks P_i release). Upon start codon recognition a series of events is triggered that includes GTP hydrolysis by eIF2 and release of eIF1/eIF2. The following 60S joining results in displacement of the remaining eIFs and is mediated by a second GTPase, eIF5B.

The complex process of initiation in eukaryotes is thought to be related to the increased structural complexity of the eukaryotic ribosome. In line with this, eIF1A binding involves r-proteins S30e and S31e (Weisser et al., 2013). First structural insights into the 43S PIC indicate that the multisubunit eIF3 interacts with several eukaryote-specific r-proteins (S1e, S26e, S27e) and r-protein extensions (S15), as well as ES6S (Hashem et al., 2013b). Moreover, the ES3S/ES6S region contributes to the binding site of eIF4G (Yu et al., 2011).

Elongation. Elongation starts with an initiator tRNA in the P-site and an empty A-site, that can accept the next A-tRNA (reviewed in Dever and Green, 2012; Voorhees and Ramakrishnan, 2013). These are delivered to the ribosome in complex with elongation factor (EF) Tu (eEF1A in eukaryotes) and GTP. A-tRNAs that are accepted during the following decoding process (see section 1.1) result in GTP hydrolysis by EF-Tu/eEF1A and its dissociation from the tRNA. In a process called accommodation the A-tRNA subsequently swings into the PTC to allow peptide bond formation (see section 1.1). The reaction results in transfer of the growing peptide chain to the newly bound tRNA in the A site and a deacylated tRNA in the P-site. EF-G (eEF2 in eukaryotes) subsequently catalyzes the translocation of the mRNA by one codon and movement of the tRNAs from the A- to the P- and from the P- to the E-site. This brings the P-tRNA back to the P-site and empties the A-site, thus preparing it for the next A-tRNA. Following GTP hydrolysis, EF-G/eEF2 leaves the ribosome. During translocation, the SSU undergoes a rotation relative to the LSU (ratcheting) (Frank and Agrawal, 2000) as well as internal movements of the head relative to the body (swiveling) (Schuwirth et al., 2005). These motions result in so called hybrid states in which tRNAs are not found in the same binding sites (A, P, E) on both subunits and tRNAs can adopt many of these intermediate states spontaneously (Fischer et al., 2010). Moreover, intrasubunit tRNA hybrid states have been observed within the SSU (Ratje et al., 2010). The ribosome alone allows movement of the tRNAs in forward and backward direction (Konevega et al., 2007; Shoji et al., 2006) and it is EF-G which guarantees directionality of the process (Frank, 2012; Frank and Gonzales, 2010). In contrast to EF-G, eEF2 is posttranslationally modified by conversion of a conserved histidine to diphthamide (Jorgensen et al., 2006). This unique modification is target for bacterial toxins, which ADP-ribosylate the diphthamide and thereby inhibit eEF2 (Dever and Green, 2012; Jorgensen et al., 2006; Mateyak and Kinzy, 2013). The exact function of diphthamide is not clear at present but it has been suggested to function during translocation by disrupting the interaction between the DC and the mRNA-tRNA duplex (Taylor et al., 2007). After translocation, deacylated tRNAs dissociate from the E-site and the essential ATPase eEF3 facilitates this clearance process in yeast (Andersen et al., 2006; Triana-Alonso et al., 1995). In addition to the classical translation factors (EF-Tu/eEF1A and EF-G/eEF2), bacterial EF-P and the orthologous eukaryotic eIF5A are also involved in the elongation cycle. They alleviate ribosome stalling by short proline-rich motifs (Doerfel et al., 2013; Gutierrez et al., 2013; Ude et al., 2013). Elongation occurs on so called polysomes, a series of ribosomes bound to the same mRNA. The relative orientation of ribosomes in polysomes with respect to their neighbors appears to be similar in bacteria and eukaryotes despite the structural differences of the 70S and 80S particles (Brandt et al., 2009; 2010). The elongation cycle continues until an mRNA stop codon is encountered in the A-site, which triggers the termination phase of translation.

Termination and recycling. During the termination phase in bacteria one of the two class-I release factors (RF) 1 or RF2 recognize a stop codon in the ribosomal A-site (RF1 recognizes UAG/UAA; RF2 recognizes UGA/UAA). These factors catalyze the hydrolysis of the P-tRNA ester bond (and thereby peptide release) by inserting a universally conserved GGQ motif into the PTC, which specifically selects water as a nucleophile in the reaction (reviewed in Klaholz, 2011; Petry et al., 2008). In the next step, RF1/RF2 are removed from the ribosome by the class-II RF3 (a GTPase) (Freistroffer et al., 1997; Zavialov et al., 2001). The following recycling phase involves the ribo-

some recycling factor (RRF) and EF-G (Zavialov et al., 2005). It results in splitting of the ribosomal subunits (along with release of mRNA and deacylated tRNA) and thus prepares them for the next round of translation. In eukaryotes a single eRF1 is used for recognition of all three stop codons. Despite being unrelated to bacterial RF1 and RF2, class-I eRF1 contains the conserved GGQ motif, which catalyzes hydrolysis of the P-tRNA. Likewise, class-II eRF3 (a GTPase) is unrelated to bacterial RF3. In contrast to RF3, it appears to ensure efficient P-tRNA hydrolysis by class-I release factors, rather than their dissociation as in the bacterial system (Alkalaeva et al., 2006). The eukaryotic recycling steps involve the essential ABCE1 ATPase (Pisarev et al., 2010) and recycling and re-initiation are tightly coupled in eukaryotes (reviewed in Dever and Green, 2012; Jackson et al., 2012).

2 Aims of this Work

Despite the existence of models of the LSU from the two archaeal species *H. marismortui* (Ban et al., 2000) and *Methanobacterium thermoautotrophicus* (Greber et al., 2012a), structural knowledge about archaeal ribosomes is still incomplete. Both organisms belong to late branching groups of the archaeal domain of life that have experienced r-protein loss during evolution (Desmond et al., 2011; Lecompte et al., 2002; Yutin et al., 2012). Moreover, no SSU model of any archaeal organism is presently available. To tackle both information gaps simultaneously, we chose a *Pyrococcus furiosus* 70S ribosome cryo-EM reconstruction at 6.6 Å as starting point to systematically model the r-protein and rRNA components (Paper 1, section 3.1). The model will provide insights into the complete archaeal ribosome architecture and given the intermediate complexity of archaeal ribosomes in comparison to the bacterial and eukaryote ones, it might also be suitable to elucidate principles of eukaryotic ribosome evolution.

For decades the eukaryotic 80S ribosome structure remained challenging for crystallography and with its technical progress, cryo-EM appeared to be a promising alternative to gain structural insights at high resolution. Parts of this dissertation aimed at providing the first complete structural models of the eukaryotic ribosome based on cryo-EM reconstructions of the yeast and wheat germ particles at 6.1 and 5.5 Å resolution, respectively (Papers 2 and 3, sections 3.2 and 3.3). First ribosome crystal structures were reported from the lower eukaryotes *S. cerevisiae* and *T. thermophila* only after publication of our results (Ben-Shem et al., 2010; 2011; Klinge et al., 2011; Rabl et al., 2011). These structures at atomic detail call for a critical assessment of the cryo-EM based models and allow an evaluation of the reliability of our modeling efforts above 5 Å resolution. Conclusions of the comparison are included in the discussion section of this thesis.

With the atomic models of lower eukaryotic ribosomes in hand it became increasingly important to extend our structural knowledge to higher eukaryotes such as humans. The substantially expanded rRNA components of these ribosomes, which are not found in unicellular eukaryotes, are likely to contribute to additional functionality in the context of the more complex metazoan cell. Parts of this thesis aimed at obtaining a cryo-EM reconstruction of the human 80S ribosome at the highest possible resolution to allow model building with great accuracy (Paper 4, section 3.4). Moreover, the human 80S structure should be supplemented with a model of the *D. melanogaster* 80S structure to provide a link to lower eukaryotic ribosomes. The intermediate size of *Drosophila* rRNA in comparison to yeast and human suggests that novel architectural rRNA features might have evolved gradually in metazoans. The human and *Drosophila* ribosome models would not only constitute the basis for future genetic, biochemical and structural studies but also complement the available 80S structures from lower eukaryotes to give a more general view on the structural evolution of the eukaryotic ribosome.

3 Cumulative Thesis: Summary of Published Results

3.1 Paper 1: Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution

Jean-Paul Armache*, Andreas. M. Anger*, Viter Márquez, Sibylle Franckenberg, Thomas Fröhlich, Elizabeth Villa, Otto Berninghausen, Michael Thomm, Georg J. Arnold, Roland Beckmann and Daniel N. Wilson

Nucleic Acids Res., 41(2), 1284-1293.
* These authors contributed equally to this work.

Genomic analysis indicates that archaeal ribosomes are of intermediate complexity compared to bacteria and eukaryotes and that the Euryarchaeota phylum has lost many r-protein families during its evolution. This is most substantial in the late branching linages such as *Halobacteria* (Desmond et al., 2011; Lecompte et al., 2002; Yutin et al., 2012). At present, only molecular models of two archaeal LSUs from *H. marismortui* and *M. thermoautotrophicus* are available. The *Haloarcula* 50S crystal structure includes the 5S and 23S rRNAs, together with 27 r-proteins (Ban et al., 2000). Additional information about 5 archaeal LSU proteins as well as some ES structures that are not present in *Halobacteria* came with the more recent cryo-EM structure of the *M. thermoautotrophicus* 50S subunit (Greber et al., 2012a). However, both organisms belong to the *Euryarchaeota* phylum and have experienced r-protein loss during evolution. Furthermore, no structural information on the archaeal SSU is presently available. In order to fill this gap, the complete molecular model of an archaeal 70S ribosome from *P. furiosus* was built based on a 6.6 Å cryo-EM structure (Becker et al., 2012). This model together with additional two-dimensional (2D) PAGE and mass spectrometry (MS) analysis of ribosomal subunits from the very closely related *Thermococcus kodakaraensis*, coupled with low-resolution cryo-EM reconstructions from various archaeal species reveal a promiscuous behaviour of r-proteins in archaea.

The molecular model of the entire *P. furiosus* rRNA shows that the majority of the present VR and ES adopt conformations remarkably similar to the equivalent regions in the eukaryotic ribosome. The model also includes the complete set of 64 (25 in SSU; 39 in LSU) r-proteins. Thus, in addition to 10 archaea/eukaryote specific r-proteins from the SSU the work presents models for L33e and L41e, which are absent in the genomes of *H. marismortui* and *M. thermoautotrophicus*. After fitting all structures in the electron density, four regions of additional protein density remained unaccounted for. 2D-PAGE and MS analysis of ribosomal subunits from *T. kodakaraensis* showed that the LSU protein L8e is also present in the SSU sample. Moreover, L8e binds characteristic kink-turn (KT) motifs in RNA and a systematic search for similar KT-motifs in the rRNA model identified two KTs in the direct vicinity of unassigned densities. Based on these results, two of the un-interpreted electron densities were filled with extra copies of L8e (termed L8e(2) and L8e(S)). Like L8e was found on the SSU, 2D-PAGE and MS identified S24e as being present on the LSU and this provided the basis for placing an additional S24e (S24e(L)). Finally, a second copy of L14e (L14e(2)) could be unambiguously fit in the last remaining density on the LSU. Taken together, three binding sites for L8e (L8e(1), L8e(2) and L8e(S)) as well as two binding sites for L14e (L14e(1) and L14e(2)) and S24e (S24e and S24e(L)) were identified in the *P. furiosus* 70S ribosome. Inspection of all available archaeal cryo-EM structures indicates that S24e(L) is specific for the

Thermococcaceae family and that L14e(2) is ubiquitously present in the archaeal phylogeny. Furthermore, results from a KT search across various archaeal species to identify additional binding sites for L8e correlate perfectly with the cryo-EM analysis. It can be suggested that L8e(S) is present in all archaea and that L8e(2) is predominantly lost in the late branching Euryarchaeota.

3.2 Paper 2: Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution

Jean-Paul Armache*, Alexander Jarasch*, Andreas M. Anger*, Elizabeth Villa, Thomas Becker, Shashi Bhushan, Fabrice Jossinet, Michael Habeck, Gülcin Dindar, Sibylle Franckenberg, Viter Márquez, Thorsten Mielke, Michael Thomm, Otto Berninghausen, Birgitta Beatrix, Johannes Söding, Eric Westhof, Daniel N. Wilson and Roland Beckmann

Proc. Natl. Acad. Sci. USA, *107*(46), 19748–19753.
* These authors contributed equally to this work.

This work dates back to 2010, a time when no crystal structure of the eukaryotic ribosome was available and structural knowledge was limited to medium to low-resolution cryo-EM reconstructions. Initial core models for the yeast 80S ribosome were built by docking rRNA structures and r-protein homology models based on bacterial/archaeal crystal structures (Ban et al., 2000; Wimberly et al., 2000) into a cryo-EM map at 15 Å resolution (Spahn et al., 2001). Subsequent extensions of these molecular models to include eukaryote-specific elements were based on cryo-EM structures of fungal and dog ribosomes (Chandramouli et al., 2008; Taylor et al., 2009). However, due to the modest resolution of about 9 Å, the completeness and accuracy of these models are also limited.

In an effort to provide the first complete molecular model of a eukaryotic ribosome, the cryo-EM structure of a wheat germ (*Triticum aestivum*) translating 80S ribosome at 5.5 Å resolution was determined and used to systematically model ~98% of the rRNA. Moreover, the wheat germ model provided the starting point for an rRNA model of the *S. cerevisiae* 80S ribosome based on a previously published cryo-EM map at 6.1 Å (Becker et al., 2009). The *T. aestivum* cryo-EM dataset contained 1,362,920 particles after rigorous *in silico* sorting for the presence of P-tRNA to increase conformational homogeneity. At time of publication it was the best-resolved electron density map of a eukaryotic ribosome. The majority (~65%) of the wheat germ rRNA was created based on homology to bacterial and archaeal crystal structures, while the remaining parts (1,903 nts) were modeled *de novo*, guided by secondary structure predictions and features of the electron density map. This strategy only left out 116 of 5,485 nts, which are mainly parts of single-stranded linkers and could not be modeled due to unreliable predictions and ambiguous electron density. The final model contained all VR and ES and allowed a complete description of the eukaryote-specific rRNA architecture. Analysis of the structure revealed a direct interaction between ES3S and ES6S via base pairing and showed that a helix of ES7L (ES7L-A) is stabilized by eukaryote-specific r-protein L28e. The latter discovery was possible due to the absence of L28e in yeast (Lecompte et al., 2002). Consequently, ES7L-A was found to be highly flexible in this organism. Wheat germ ES7L also contains a three-way junction formed by helices C, D and E, which is not present in *S. cerevisiae*. The N-terminus of r-protein L6e, which is shorter in yeast, appears to insert through this ES7L three-way junction and thus forms a novel type of RNA-protein interaction that was described for the first time in this publication. The paper also gives insights into the dynamic behaviour of ES27L. In addition to the known ES27L-in and -out conformations

(Beckmann et al., 2001), analysis of the yeast structure reveals a new intermediate position, termed ES27-int. The ES27L-int and ES27L-in conformations appear to be stabilized by interactions with the eukaryote-specific r-proteins L38e and L34e, respectively (Note that L34e was localized incorrectly and the ES27L-in stabilizing protein turned out to be L27e (Ben-Shem et al., 2011; Klinge et al., 2011)).

Today, the results presented in this publication have to be seen in light of the crystal structures of the eukaryotic ribosome and a critical assessment of the cryo-EM based rRNA models is included in the discussion section of this thesis.

3.3 Paper 3: Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome

Jean-Paul Armache*, Alexander Jarasch*, Andreas M. Anger*, Elizabeth Villa, Thomas Becker, Shashi Bhushan, Fabrice Jossinet, Michael Habeck, Gülcin Dindar, Sibylle Franckenberg, Viter Márquez, Thorsten Mielke, Michael Thomm, Otto Berninghausen, Birgitta Beatrix, Johannes Söding, Eric Westhof, Daniel N. Wilson and Roland Beckmann

Proc. Natl. Acad. Sci. USA, 107(46), 19754–19759.
* These authors contributed equally to this work.

Like paper 2, this work from 2010 dates back to a time when no crystal structure of the eukaryotic ribosome was published. Compared to 54 r-proteins in bacteria, the eukaryotic ribosome contains ~80 r-proteins. Information on the localization of proteins within the 80S ribosome mainly has come from immuno-EM and crosslinking studies (Gross et al., 1983; Lutsch et al., 1990; Marion and Marion, 1987). Docking of bacterial/archaeal crystal structures into a 15 Å cryo-EM map of *S. cerevisiae* later identified the location of a total of 43 eukaryotic r-proteins which have bacterial or archaeal homologs (Spahn et al., 2001) (Note that position of L39e is known from the 50S crystal structure of *H. marismortui* (Ban et al., 2000), but was not included in the initial yeast model from Spahn et al.). In addition to this, the localization of r-proteins RACK1, S19e and L30e have been elucidated more recently in cryo-EM structures of plant and fungal 80S ribosomes (Halic et al., 2005; Sengupta et al., 2004; Taylor et al., 2009). This leaves a total of 30 r-proteins, excluding the stalk proteins (L10, P1 and P2), with unknown localization in the eukaryotic ribosome.

The wheat germ 80S cryo-EM structure from paper 2 together with a *S. cerevisiae* map at 6.1 Å (Becker et al., 2009) allowed the identification and modeling of 74 of the 80 r-proteins in the eukaryotic ribosome. This includes 27 r-proteins (excluding the stalk proteins L10, P1 and P2), which are not present in the bacterial or archaeal crystal structures. The r-protein models complement the rRNA structures from paper 2 to give near-complete models of the 80S ribosomes from *T. aestivum* and *S. cerevisiae*. 44 proteins were built using templates of the archaeal and bacterial crystal structures. Eukaryote-specific extensions were modeled *de novo* whenever possible using electron density and secondary structure constraints. 17 r-proteins (S4e, S17e, S19e, S24e, S27e, S28e and RACK1 for the SSU; L4e, L6e, L14e, L20e, L27e, L30e, L33e, L10, P1 and P2 for the LSU) were modeled using available nuclear magnetic resonance (NMR) spectroscopy or crystal structures of free r-proteins. Homology models for 6 r-proteins (S25e, L22e, L29e, L34e, L36e and L38e) are based on similarity to non-ribosomal proteins with known structure. The remaining 7 r-proteins (S7e, S21e, S26e, S30e, L13e, L28e and L41e) were tentatively modeled *ab initio*. 6 small

r-proteins of the SSU (S1e, S6e, S8e, S10e, S12e and S31e) could not be localized and were therefore not modeled. Primary basis for the localization of r-proteins was the excellent agreement of protein folds of the crystal structures and homology models with the electron density features at 5.5 Å resolution. For *ab initio* modeled r-proteins previously published results from crosslinking and immuno-EM studies were additionally taken into consideration. L38e was localized on the basis of an 80S cryo-EM reconstruction from a yeast strain lacking this non-essential protein. L28e and L34e were positioned based on comparison of ribosome structures of *S. cerevisiae* (lacking L28e) with *T. aestivum* and *H. marismortui* (lacking L34e) with *P. furiosus*, respectively. The models presented in this work reveal, that in contrast to bacteria, several eukaryote-specific r-proteins and r-protein extensions reach into functional sites of the conserved ribosome core. These include the DC (S4, S30e), the tRNA binding sites (S25e) and the PTC (L16) (Note that S30e was localized incorrectly and the protein tail reaching into the DC actually belongs to S31e (Rabl et al., 2011)). Moreover, S26e and S28e were found to constitute part of the mRNA exit site on the eukaryotic 40S subunit and previously unknown r-protein interaction partners of yeast eEF3 (Andersen et al., 2006) were identified as S19e and S25e. The 80S models published in papers 2 and 3 uncover rRNA and r-protein coevolution as a prominent theme in the architecture of eukaryotic ribosomes. This is exemplified by the intertwined structure on the back of the LSU, that is formed by ES7L and ES39L together with the eukaryote-specific r-proteins L6e, L14e, L18ae, L28e and L33e.

Similar to the rRNA models presented in paper 2, the localization of eukaryote specific r-proteins presented in this work are thoroughly compared to the results obtained by crystallography in the discussion section of this thesis.

3.4 Paper 4: Structures of the human and *Drosophila* 80S ribosome

Andreas M. Anger*, Jean-Paul Armache*, Otto Berninghausen, Michael Habeck, Marion Subklewe, Daniel N. Wilson and Roland Beckmann

Nature, 497(7447), 80-85.
* These authors contributed equally to this work.

Crystal structures have elucidated the architecture of lower eukaryote 80S ribosomes. In contrast, the limited resolution (9 to 20 Å) of cryo-EM structures of mammalian 80S ribosomes has so far prohibited the generation of complete molecular models for these higher eukaryotes.

Cryo-EM structures of human and *D. melanogaster* 80S ribosomes were determined with average resolutions of 5.4 and 6.0 Å, respectively. However, the official resolution of 5.4 Å for the human ribosome does not reflect the true quality of the map and local resolution determination reveals numbers better than 4.8 Å for large parts of the structure, with the best-resolved areas even reaching towards 4.0 Å. The high quality electron density maps coupled with secondary structure predictions for the rRNA ES and the available crystal structures allowed to build complete molecular models of the fly and human 80S ribosome. Both ribosomes are in complex with E-site tRNA, eEF2 and Stm1-like stress proteins (SERBP1 and Vig2 in human and *Drosophila*, respectively) and the identities of the latter two factors were confirmed my MS analysis. The presence of SERBP1 and Vig2 on higher eukaryote ribosomes indicates a novel role for these proteins, analogous to Stm1 in yeast (Ben-Shem et al., 2011), in the regulation of translation in human and fly. Description of the complete protein architecture only reveals a modest increase in protein mass and extensions of r-proteins like L6e appear to be involved in stabilization of the

intertwined RNA-protein layer, whose dimensions have developed further in higher eukaryotes. In addition, the models allowed mapping of previously unknown contacts of r-proteins S30e, S31e, L10 and L11 with eEF2, as well as modeling of mammalian specific insertions in the G' domain of the elongation factor. Moreover, in the human 80S map, electron density can be observed for the unique diphthamide of eEF2 and indications of two alternative conformations of the modification are visible, one pointing towards the DC, the other contacting the Stm1-like protein SERBP1. The complete ES inventory of the *Drosophila* and human rRNA reveals several ES that are substantially expanded compared to yeast. These include ES7L, ES15L, ES27L and ES39L that contain long flexible helical insertions and extensions that protrude from the surface of the human ribosome. The outer parts of these flexible rRNA tentacles are not visible in the cryo-EM reconstruction but observable in individual electron-microscopy images. The *Drosophila* rRNA structure also reveals several species-specific variations in ES length and topology, including a helical insertion in ES6S as well as uniquely elongated ES9S and ES31L. In addition to the ES3S-ES6S pseudoknot that appears to be conserved throughout the eukaryotic domain of life, a novel base pairing interaction between ES9L and ES15L could be identified in the human ribosome. Comparison of the eEF2-bound (rotated) state of the *Drosophila* ribosome with a subpopulation that lacked eEF2 and exists in a non-rotated state reveals a surprising dynamic interplay of structural rearrangements of ES27L and ES31L. Both ES appear to be engaged in different sets of intersubunit-bridges with r-proteins S1e, S8e and S27e, depending on the state (rotated or non-rotated) of the ribosome. Based on the yeast and *Tetrahymena* ribosome crystal structures it is known that eukaryotic ES31L and ES39L contain extended single-stranded rRNA parts that are used as platforms for r-protein binding (Ben-Shem et al., 2011; Klinge et al., 2011). Additional non-helical rRNA stretches can be observed in human and *Drosophila* ES7L, ES10L, and ES15L. These are not only used for RNA-protein interactions but also establish unique RNA-RNA interactions that contribute to the stabilization of the extended ES cluster on the back of the LSU of higher eukaryotes.

Inspection of the lower eukaryote 80S structures together with the *Drosophila* and human models presented in this work reveal a layered evolution of the eukaryotic ribosome. The intertwined rRNA-protein layer observed in lower eukaryote 80S ribosomes has increased in size and complexity in higher eukaryotes. Moreover, the substantial increase in RNA mass of higher eukaryotes, particularly mammalian ribosomes has resulted in the presence of two additional RNA layers: a rigid inner layer, resulting from multiple RNA-RNA tertiary interactions, followed by a flexible outer layer, arising from helical insertions and extensions of the rRNA ES.

4 Discussion

4.1 Critical Assessment of the Cryo-EM Based Lower Eukaryote Ribosome Models and the Importance of Resolution

Shortly after finishing the *T. aestivum* and *S. cerevisiae* 80S ribosome models based on cryo-EM reconstructions at 5.5 and 6.1 Å resolution, respectively, an initial crystal structure of the yeast 80S ribosome at 4.15 Å was reported (Ben-Shem et al., 2010). This was followed by additional crystal structures of the 40S and 60S subunits of *T. thermophila* at 3.9 Å and 3.5 Å, respectively (Klinge et al., 2011; Rabl et al., 2011), as well as an improved yeast 80S structure at 3.0 Å resolution (Ben-Shem et al., 2011), which collectively provide insights into the lower eukaryotic translation apparatus in atomic detail. With the crystal structures in hand, a critical quality assessment of the cryo-EM based rRNA and r-protein models is needed to evaluate the reliability of modeling efforts above 5 Å resolution. This assessment was done by systematic comparison of our yeast model with the 3.0 Å 80S crystal structure for both rRNA and r-proteins. Although the yeast cryo-EM model is officially based on a 6.1 Å map, it benefited substantially from insights gained during modeling of the *T. aestivum* 80S at 5.5 Å. Given the high similarity of the yeast and wheat germ cryo-EM structures, the majority of the yeast model can be seen as a copy of the corresponding *T. aestivum* parts. Thus, conclusions drawn from a comparison of the yeast models (X-ray vs. cryo-EM) can also be transferred to the *T. aestivum* 80S ribosome structure.

Ribosomal proteins. On the SSU the localization of 8 eukaryote-specific proteins (S4e, S7e, S19e, S25e, S26e, S27e, S28e and RACK1) was confirmed by the yeast crystal structure (Figure 8) (Ben-Shem et al., 2011). With the positions of S19e and RACK1 already determined by previous cryo-EM studies (Sengupta et al., 2004; Taylor et al., 2009), this leaves the locations of 6 r-proteins within the SSU that were revealed in this work. Although localized correctly, S7e and S26e were modeled with different folds compared to the crystal structure. This is due to insufficient electron density connections for these proteins in combination with the lack of homologous structures being available that could be used as modeling templates. The positions of r-proteins S17e, S21e, S24e and S30e were assigned differently in the cryo-EM model compared to the crystal structure. Three of these wrongly localized proteins were placed in the beak (S17e, S30e) and foot (S24e) regions of the SSU. Both parts of the structure are known to exhibit more ambiguous electron density due to the inherent flexibility of the 40S subunit. Accordingly, r-proteins S6e, S8e, S10e, S12e and S31e, that could not be localized in the cryo-EM map are also situated in the beak and foot regions (Figure 8) (Ben-Shem et al., 2011). Positioning of S21e and S24e was based on results from immuno-EM (Bommer et al., 1991) and the incorrect place of these two r-proteins in the cryo-EM model is not far off their true location revealed by the crystal structure (Figure 8). S30e has been shown to crosslink to mRNA (Bulygin et al., 2005; Takahashi et al., 2002). The cryo-EM based model of this r-protein contains an extension that reaches into the DC of the 40S subunit and thus could provide the basis for the crosslink result. However, it turned out that the correct S30e position (which is in direct vicinity to mRNA at the 40S shoulder) was used for wrong placement of eukaryote-specific extensions of S4 in the cryo-EM density. A comparable scenario led to misplacement of S17e. Its true position was interpreted as an extension of S2, which in fact is disordered in the crystal structure.

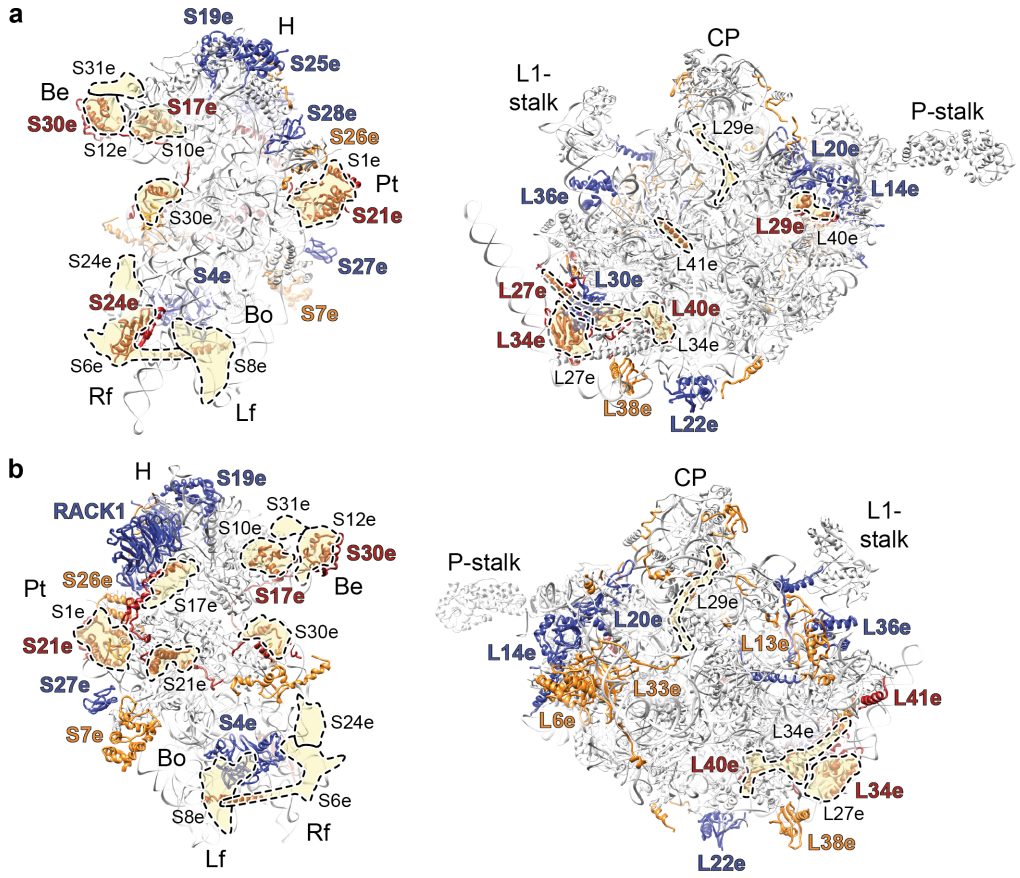


Figure 8 | Localization of eukaryote-specific r-proteins in the cryo-EM based yeast model. a, **b**, Interface (**a**) and solvent (**b**) view of the yeast 40S (left) and 60S (right) ribosome subunit from (Armache et al., 2010a; 2010b). Correctly localized proteins are colored in blue. Protein parts that were assigned properly but modeled with incorrect folds are highlighted in orange, while wrongly placed models are shown in red. As reference, r-protein positions from the *S. cerevisiae* 80S crystal structure (Ben-Shem et al., 2011) are indicated in pale yellow with dashed outline.

Within the LSU, positions of 10 eukaryote-specific r-proteins (L6e, L13e, L14e, L20e, L22e, L30e, L33e, L36e, L38e and L28e) were confirmed by the yeast and *Tetrahymena* crystal structures (Ben-Shem et al., 2011; Klinge et al., 2011). Note that L28e is missing in yeast but localization and fold of the protein in wheat germ is in agreement with the 60S crystal structure from *T. thermophila* (Klinge et al., 2011). Localization of L30e has already been reported previously (Halic et al., 2005), which leaves the positions of 9 eukaryote-specific r-proteins revealed in this work. Although localized correctly, L6e, L13e, L33e and L38e were modeled with different folds in comparison to the proteins in the crystal structure. Due to the absence of reliable homologous template structures, L13e had to be entirely modeled *de novo*. The resulting model suffered from ambiguous densities for several loop regions that collectively led to wrong connectivity of the remaining α-helical parts of the protein. In contrast to this, L6e was based on a reliable template structure, which includes the central SH3-like β-barrel fold of the protein. However, the template only covers a fraction of the protein and remaining parts of the L6e density were interpreted *de novo*, which appears to be error-prone above 5 Å resolution. In principle the cryo-EM based L33e model shows the correct β-barrel fold but was placed in density with a 180° rotation in comparison to the crystal structure. At 5.5 Å resolution individual β-sheets appear as smooth surfaces and individual strands are not separated. In the absence of additional flanking folds, which is the case for L33e, a β-barrel model can be placed in a 5.5 Å density in two orientations

(related by a 180° rotation), giving rise to basically equally satisfying interpretations. L38e contains a central β-sheet paced against two α-helices. This fold resembles a KH domain when observed in a 5.5 Å electron density map and accordingly the L38e model was based on a KH domain template. Despite this apparent similarity, secondary structure elements of the L38e crystal structure are connected differently and the protein does not belong to the KH family. Lack of β-strand separation and ambiguous density for loop regions at the given cryo-EM resolution prevented recognition of these discrepancies between template and experimental density, resulting in a wrong L38e fold in comparison to the crystal structure. Positions of L27e, L29e, L34e, L40e and L41e are not in agreement with the crystal structure (Figure 8). Localization of L34e was based on a comparison of the *H. marismortui* 50S crystal structure (Ban et al., 2000) with a cryo-EM reconstruction of the *P. furiosus* 70S ribosome at 10 Å resolution that was available at the time (Armache et al., 2010b). The logic behind this strategy was that L34e is present in the *P. furiosus*, but not the *H. marismortui* genome (Lecompte et al., 2002; Yutin et al., 2012). Position of L34e was not confirmed by the yeast 80S crystal structure, but surprisingly turned out to be occupied by L27e. Notably, L27e is absent in both *P. furiosus* and *H. marismortui* (Lecompte et al., 2002; Yutin et al., 2012), raising the question which protein instead of L27e binds at this position in the *P. furiosus* map. The answer was later given by our modeling efforts on the *P. furiosus* ribosome presented in this work. The eukaryotic L27e position appears to be used by a second copy of the promiscuous archaeal protein L14e in *Pyrococcus* (Armache et al., 2013), giving rise to the electron density that was wrongly assigned to L34e. Placement of L27e was guided by crosslink data that points to a location near L34e (Marion and Marion, 1987), as well as comparison of the eukaryotic 80S maps (*S. cerevisiae* and *T. aestivum*) with the *H. marismortui* 50S crystal structure (Ban et al., 2000) and a 10 Å cryo-EM reconstruction of the *P. furiosus* 70S ribosome (Armache et al., 2010b) (both archaeal species lack L27e). However, the yeast crystal structure revealed that L34e was mistaken for L27e in the cryo-EM based model. As already mentioned, L34e is present in the *P. furiosus* but not the *H. marismortui* genome (Lecompte et al., 2002; Yutin et al., 2012), which brings up the question why the presence of L34e was missed during inspection of the 10 Å *P. furiosus* map. Systematic modeling of the *P. furiosus* 70S ribosome at a higher resolution later showed that the majority of L34e is buried within the rRNA and only a small helical stretch of 17 aa is visible on the surface of the archaeal ribosome (Armache et al. 2013), which is to small to be unambiguously recognized at 10 Å. In eukaryotes this helix is enlarged and clearly visible in the yeast and wheat germ maps. Positioning of L29e in a small pocket under the P-stalk was based on the observation of stalk rearrangements in an 80S ribosome cryo-EM map from a yeast strain lacking the gene for L29e (ΔL29e). Moreover, the cryo-EM based position is close to L16, which exhibits synthetic lethality with L29e in yeast (DeLabre et al., 2002). However, it turned out that the ΔL29e 80S reconstruction was misleading and the reason for the observed stalk rearrangement remains unclear. In the crystal structure, L29e is found at a position in direct vicinity to L16 near the CP. Correct localization of L29e in the cryo-EM map was further complicated by the facts, that it is a small r-protein (59 aa) and adopts an extended conformation. Likewise, the remaining mispositioned L40e and L41e are the smallest r-proteins (52 and 25 aa, respectively) and due to their size lack major recognizable tertiary structure features which could have guided their placement in the electron density.

Taken together, of the 30 r-proteins with unknown localization in the eukaryotic ribosome, 15 (SSU: S4e, S7e, S25e, S26e, S27e and S28e; LSU: L6e, L13e, L14e, L20e, L22e, L28e, L33e, L36e, and L38e) were correctly placed in the cryo-EM density map. Another 9 (SSU: S17e, S21e, S24e and S30e; LSU: L27e, L29e, L34e, L40e and L41e) were positioned incorrectly and 6 r-proteins of the SSU (S1e, S6e, S8e, S10e, S12e and S31e) could not be localized and therefore were not modeled.

Mislocalized proteins primarily resulted from bad electron density due to flexibility of the SSU or their small size in combination with a lack of prominent structural features in the case of the LSU. Proteins with wrong folds lacked reliable templates in most cases, which indicates that *de novo* protein modeling at 5.5 Å resolution is ambitious and the quality of the resulting structure very likely suffers from ambiguous and misleading density connections.

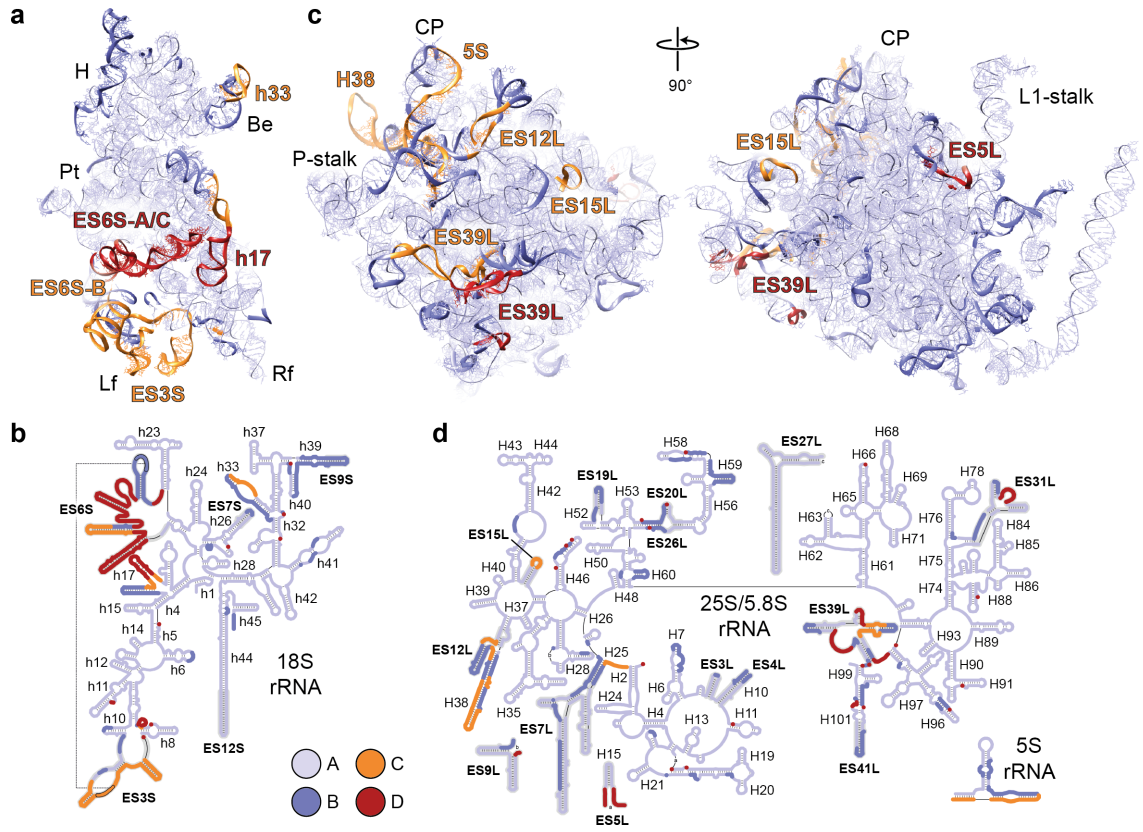


Figure 9 | Assessment of the cryo-EM based yeast rRNA model. (**a**,**b**) structure (**a**) and secondary structure diagram (**b**) of the yeast rRNA from (Armache et al., 2010a) with the quality of the model in comparison to the crystal structure from *S. cerevisiae* (Ben-Shem et al., 2011) indicated in four categories (A-D) colored in light blue (A), violet (B), orange (C) and red (D). For details on the classification criteria see main text. (**c**,**d**) structure (**c**) and secondary structure diagram (**d**) of the yeast rRNA from (Armache et al., 2010a) with the quality of the model indicated as in (a,b).

Ribosomal RNA. Systematic comparison of the *S. cerevisiae* rRNA models allowed classification in four categories (A to D) (Figure 9 and Table 1). Category A includes all parts of the cryo-EM based rRNA model that are identical or show a maximal frameshift of 0.5 nt compared to the crystal structure. Category B contains parts with shifts between 0.5 and 2 nts. Category C summarizes portions that show frameshifts of more than 2 nts but in principle follow the path of the reference structure. Category D indicates not modeled parts or stretches with wrong path and/or connectivity. 90% of the cryo-EM based model falls into categories A and B (Table 1). Notably, this vast majority of the structure does not only contain the conserved rRNA core but also many *de novo* modeled parts of the SSU (ES9S, ES12S, h6, h16 and h41) (Figure 9a,b) and LSU (ES3L, ES4L, ES7L, ES9L, ES10L, the majority of ES12L, ES19L, ES20L, ES26L, ES31L, ES41L, H16-H18 and H59) (Figure 9c,d), underscoring the high degree of reliability of rRNA modeling using density maps at 5.5 Å resolution (Figure 9). It has to be noted that the classification presented here includes ES7L-

A, ES27L and the L1-stalk (H77-H78) in the best category (A). These parts of the rRNA are not included in the *S. cerevisiae* crystal structure due to their high flexibility (Ben-Shem et al., 2011). A correction, by ignoring these stretches in the quality assessment, reduces the percentage of the cryo-EM based rRNA model contained in categories A and B only marginally from 90 to 89%. Half of the remaining model (5%) falls into category C (Table 1). This includes parts of ES3S, ES12L, ES15L, ES39L, h33, H38 and 5S rRNA (Figure 9). Despite more than 2 nts frameshift observed in these models, they still follow the path of the crystal structure, and thus deliver a correct picture of the principal rRNA architecture. 122 nts (2%) of rRNA that were not modeled because of ambiguous electron density and 158 nts (3%) of modeled parts with connectivity errors or large deviation from the crystal structure are combined in category D (Table 1). The latter fraction of the model includes h17, as well as parts of ES6S, ES5L and ES39L (Figure 9). Reasons for the modeling failure of these parts will be discussed below. In general, the cryo-EM based rRNA model of the LSU is in in better agreement with the crystal structure than the SSU model (83% vs. 70% in category A, respectively) (Table 1). The overall quality of the SSU rRNA suffers from the inherent flexibility of the subunit, which results in less resolved electron densities particularly in the beak and foot regions. Accordingly, h33 and ES3S, which constitute large parts of these structures, contain severe frameshifts of more than 2 nts in comparison to the crystal structure (category C) (Figure 9a,b). The degree of correctness for the rRNA model correlates well with results from the r-protein assessment, where the majority of errors are also found within the beak and foot regions of the SSU (Figure 8).

Table 1 | Assessment of the yeast rRNA model.

		Category A		Category B		Category C		Category D	
rRNA	length (nts)	nts	%	nts	%	nts	%	nts	%
18S	1800	1258	70	225	13	131	7	186	10
5S	118	46	39	40	34	32	27	0	0
5.8S	158	127	80	30	19	0	0	1	1
25S	3396	2835	83	375	11	93	3	93	3
Total	5472	4266	78	670	12	256	5	280	5

Note that a different 5S rRNA sequence of 121 nts in length was used in (Ben-Shem et al., 2011).

The majority of the *de novo* modeled rRNA parts are in good agreement with the crystal structure. Particularly successful examples include ES7L, ES8L, ES9L, ES20L, ES26L and ES31L, all with a root mean square deviation (RMSD) below 1.4 Å between the two models (Figure 10). All of these rRNA regions are dominated by defined secondary structures in the form of standard RNA helices (ES7L-A to C, H28 to H31) (Figure 10a,b) or contain only short (1 to 5 nts) non-helical stretches (ES20L and ES26L) (Figure 10c). A large content of helical elements connected by short linkers allows reliable secondary structure predictions, which in turn provide the basis for *de novo* modeling. Moreover, RNA helices can be readily recognized as ribbon-like densities even in maps with moderate resolution. In contrast to this, resolution becomes limited for extended (>10 nts) single-stranded parts such as the connection between ES31L helices A and B (Figure 10d). This part of the structure could not be modeled reliably and was left out in the final rRNA model.

Less successful rRNA modeling examples include ES6S, ES5L and ES39L (Figure 11). The failure to correctly model the KT of ES5L is difficult to explain. The structural motif is also present in the 50S crystal structure from *H. marismortui*, where it forms the binding site of the archaea-/eukaryote-specific r-protein L8e (Ban et al., 2000) and is found to be nearly identical in the yeast 80S crystal structure (Ben-Shem et al., 2011). It seems that the presence of the KT motif was overlooked during the initial structural alignments to generate the rRNA core. Subsequent *de novo* modeling efforts for this region also failed to identify the motif due to ambiguous density (Figure 11a).

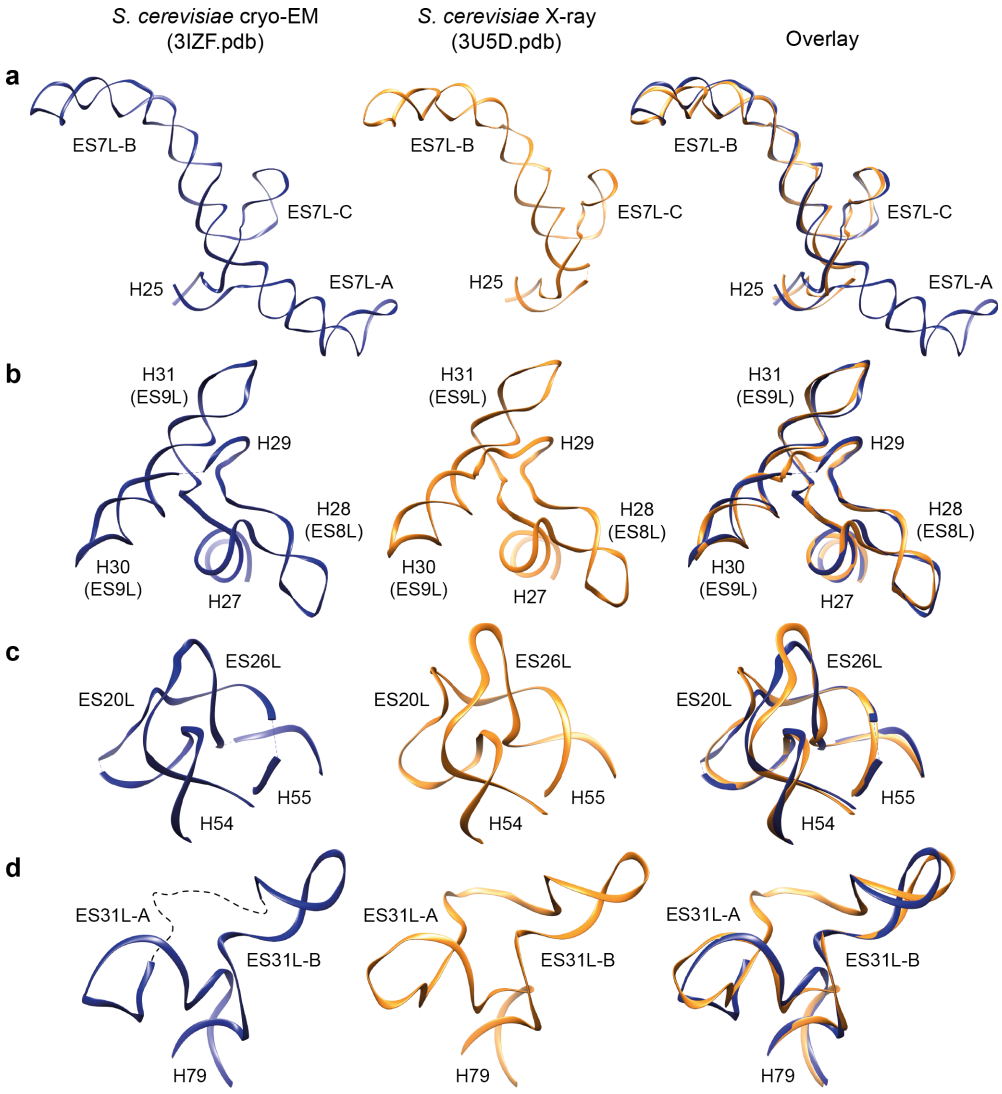


Figure 10 | rRNA modeling highlights. (**a**-**d**) Comparison of rRNA parts from the cryo-EM based yeast 80S model (Armache et al., 2010a) (left, blue) with the *S. cerevisiae* ribosome crystal structure (Ben-Shem et al., 2011) (middle, orange), together with an overlay of both (right). (**a**) ES7L with RMSD of 1.38 Å between the two models. Note that ES7L-A is not included in the crystal structure due to its high flexibility. (**b**) ES8L/ES9L, RMSD 1.25 Å. (**c**) ES20L/ES26L, RMSD 1.32 Å. (**d**) ES31L, RMSD 1.31 Å. ES numbering follows the extended definition from (Anger et al., 2013).

ES31L and ES39L contain long single-stranded rRNA linkers that were problematic during model building (Figure 11b,c). While the non-helical stretch was left out in the final ES31L model, a complete solution for ES39L was included but turned out to differ significantly in comparison to the crystal structure, even showing wrong connectivity between the helical parts of this ES. ES31L and ES39L linker densities appear fragmented and are highly ambiguous in the cryo-EM map (Figure 11b,c). In the case of ES39L the situation is further complicated by the fact that the ES is intertwined with several eukaryote-specific proteins (L6e, L14e, L20e, L33e) and r-protein extensions (L13, L22), making it difficult to assign density to r-proteins or single-stranded rRNA. Indeed, parts of a ES39L linker were positioned where an extension of L22 is located in the crystal structure (Figure 11c).

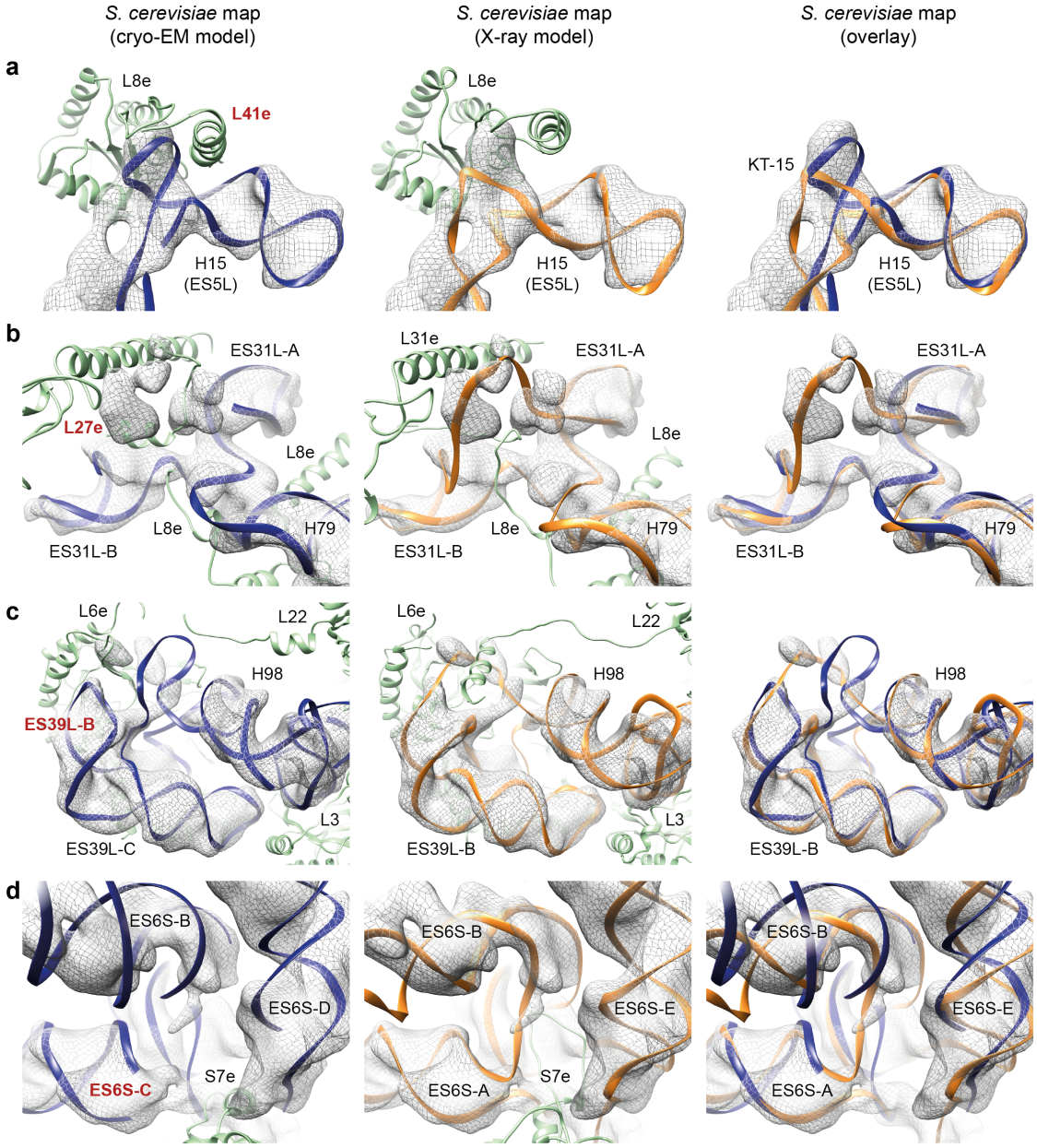


Figure 11 | Ambiguous electron density for single-stranded rRNA. (**a**-**d**) Comparison of rRNA parts from the cryo-EM based yeast 80S model (Armache et al., 2010a) (left, blue) with the *S. cerevisiae* ribosome crystal structure (Ben-Shem et al., 2011) (middle, orange), together with an overlay of both (right). RNA electron density of the yeast 80S cryo-EM reconstruction that was used for modeling is shown as grey mesh (**a**) ES5L. (**b**) ES39L. (**c**) ES31L. (**d**) ES6S. ES numbering follows the extended definition from (Anger et al., 2013).

The strategy to rely on secondary structure predictions during *de novo* modeling turned out to be problematic for ES39L. The precise position of rRNA helices, connected by long single-stranded parts, are difficult to predict in sequences via the minimum free energy because the algorithms tend to include non-helical elements in base pair interactions. In the ribosome long single-stranded rRNA parts are prevented from forming stable but functionally misfolded states by tertiary interactions with r-proteins and/or rRNA. For instance, structure-sensitive chemical probing has been utilized to show that *in vitro* transcribed ES6S folds different in comparison to ES6S in the ribosomal context (Alkemar and Nygård, 2006). In the case of ES39L, even the

combination of experimental data from structure-sensitive reagent probing with thermodynamic energy minimization did not result in a correct prediction of the single-stranded regions (Nygård et al., 2006).

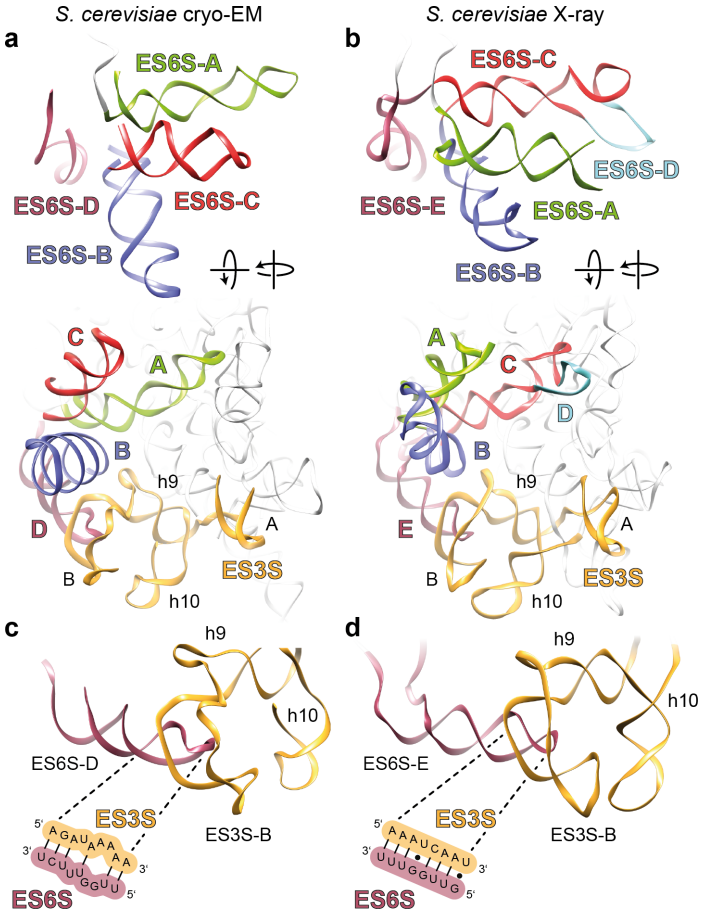


Figure 12 | Models of the ES3S-ES6S region in *S. cerevisiae*. (**a**,**b**) Comparison of ES6S from the cryo-EM based yeast 80S model (**a**) (Armache et al., 2010a) with the *S. cerevisiae* ribosome crystal structure (**b**) (Ben-Shem et al., 2011). Corresponding helices (ES6S-A to E) are colored distinctly. ES3S is shown in orange (**c**,**d**) ES3S-ES6S interaction in the cryo-EM based yeast 80S model (**c**) (Armache et al., 2010a) and in the *S. cerevisiae* ribosome crystal structure (**d**) (Ben-Shem et al., 2011).

ES6S is located in the lower part of the SSU and electron density in the wheat germ and yeast cryo-EM maps allowed clear identification of four helices (ES6S-A to D) for this ES (Figure 12a). However, the single-stranded linker regions between the four helices at the base of ES6S were not well resolved (Figure 11d). Density interpretation is further complicated by the presence of r-protein S7e, which stabilizes the base of ES6S. At the given resolution it is difficult to dissect protein and single-stranded rRNA densities in this region of the map and as a consequence, the ES6S linker regions could not be reliably modeled (Figure 11d). Without knowledge about the connectivity between the four ES6S helices their correct placement is impossible to determine solely on the electron density map and additional information is required. Position of ES6S-B was guided by the observation of this helix being branched in a preliminary cryo-EM structure of the *Drosophila* 80S ribosome and the corresponding part in the yeast and wheat germ 18S rRNA could readily be determined by sequence alignments. One ES6S helix (ES6S-D in our model; Figure 12a) runs to the bottom of the 40S subunit where it appears to interact with ES3S and this interaction has been suggested to occur via base-pairing in a pseudoknot structure formed by the two ES (Alkemar and Nygård, 2003). However, the ES3S-ES6S pseudoknot was not included in a previous cryo-EM based models of the 80S ribosome (Chandramouli et al., 2008; Taylor et al., 2009). To allow modeling of the ES3S-ES6S base-pairing interaction, parts of ES3S had to reach from the right side to the left foot region of the subunit where ES6S is located (Figure 12a,c). Such an rRNA

path is consistent with conclusions drawn from inspection of a mammalian ribosome cryo-EM structure, which shows an elongated left foot, but no substantial extension on the right side of the SSU (Chandramouli et al., 2008). On the sequence level mammalian ES3S is extended in comparison to yeast, while the size of ES6S remains rather constant and thus can not account for the enlarged left foot (Cannone et al., 2002). The base-pairing nucleotides between ES3S and ES6S in the wheat germ cryo-EM based model (not shown) and the yeast crystal structure (Figure 12d) are in perfect agreement with predictions for the interaction (Alkemar and Nygård, 2006). In the yeast cryo-EM based model however, the ES3S-ES6S pseudoknot has been modeled in a way that involves different stretches of nucleotide compared to the crystal structure (Figure 12c,d). This is the direct result from substantial frameshifts in the yeast ES3S model (Figure 9a,b), that are not found to the same extend in the wheat germ structure (not shown). Taken together, additional information beyond the electron density allowed reliable positioning of the hybrid forming helix of ES6S and ES6S-B even without knowledge of the linkers connecting them. Both helix locations, together with the pseudoknot between ES3S and ES6S were later confirmed by the crystal structure (Figure 12b,d).

In contrast to this, positions of ES6S-A and C in the cryo-EM based model are not in agreement with the crystal structure (Figure 12a,b). Placement of these helices had the least supporting data and was largely based on speculation. Bacterial 16S rRNA is tolerant towards insertions 3' of h21 (Yokoyama and Suzuki, 2008), which allows the assumption that ES6S might have originated from such an insertion during ribosomal evolution and that the sequences near the 5' end of ES6S correspond to h21 in the eukaryotic ribosome. Following this argument, we placed ES6S-A helix as the h21 corresponding part. ES6S-C then ended up in the last remaining density. It has to be pointed out that our final placement of the ES6S-A and C helices only relies on weak reasoning and counterarguments also exist. For instance, structure-sensitive chemical probing indicates that the ES6S-C (ES6S-D in X-ray) part is less modified than other portions of ES6S and thus seems to be less accessible in the ribosome (Alkemar and Nygård, 2006). This argues against our placement, in which the ES6S-C helix is protruding from the surface of the 40S subunit (Figure 12a). All secondary structure predictions for ES6S missed the existence of an extra helix (ES6S-D in the crystal structure) that creates a three-way junction close to the tip of ES6S-C (Figure 12b). Electron density corresponding to the short extra helix could be observed in both the yeast and wheat germ cryo-EM structures. However, due to the limited resolution of the maps, which especially affects the surface of the SSU, it was not possible to clearly decide whether the density belongs to RNA or protein. Note that the helix naming differs between the cryo-EM based ES6S model and the crystal structure, where the hybrid forming helix is labeled ES6S-E due to the presence of the short additional helix ES6S-D (Figure 12b).

4.2 Higher Eukaryotic Ribosome Models

The *H. sapiens* 80S ribosome model presented in this thesis is based on a cryo-EM reconstruction with an average resolution of 5.4 Å. Although this number is basically identical to the 5.5 Å of the *T. aestivum* structure, it underestimates the true quality of the human density map. This is due to the presence of very flexible rRNA tentacles protruding from the surface the human ribosome. These flexible parts drastically influence the overall resolution and in fact, determination of the local resolution indicates better than 4.8 Å for large parts of the structure with the best-resolved areas even reaching towards 4.0 Å (Anger et al., 2013). At resolution below 5 Å, extended single-stranded rRNA linkers become traceable in the electron density (Figure 13), which allows overcoming the major obstacle in the generation of reliable *de novo* models for complicated rRNA

arrangements. In retrospect, the ability to follow the electron density during *de novo* modeling of the non-helical parts of ES15L turned out to be the only working strategy resulting in a complete model of this rRNA segment (Figure 13a). Due to the very long (>20 nts) single-stranded parts of ES15L, structure prediction attempts failed to produce a solution that is close to the folding of the ES in the ribosomal context (Larsson and Nygård, 2001).

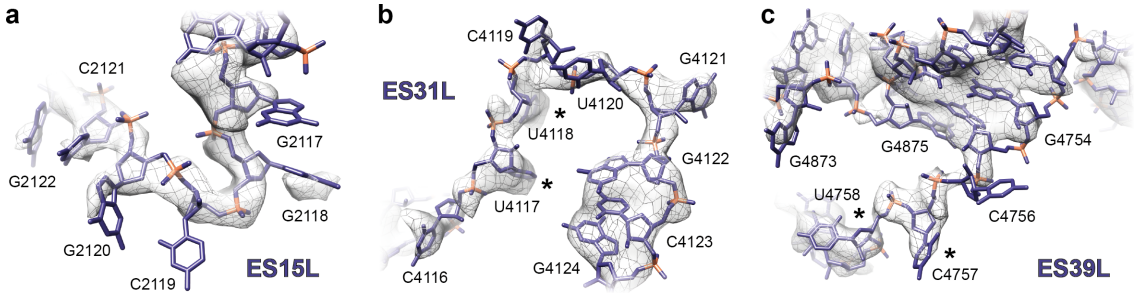


Figure 13 | Single-stranded linkers in human rRNA. (**a**-**c**) non-helical rRNA segments from human ES15L (**a**), ES31L (**b**) and ES39L (**c**). Backbone phosphates are highlighted in orange. Bases involved in stacking interactions with r-protein side chains are indicated with asterisks.

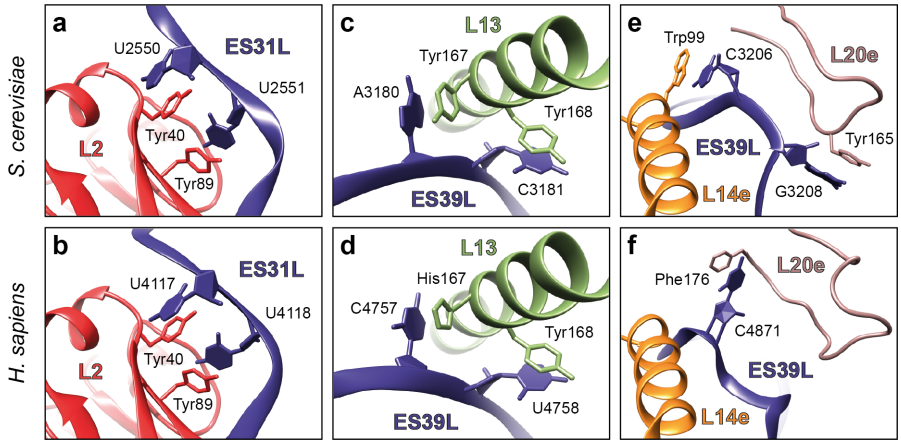


Figure 14 | Single-stranded rRNA as platform for r-protein binding. (**a**,**b**) Interaction of L2 with ES31L in *S. cerevisiae* (**a**) (Ben-Shem et al., 2011) and *H. sapiens* (**b**) (Anger et al., 2013). (**c**,**d**) Binding of ES39L to r-protein L13 in yeast (**c**) and human (**d**). (**e**,**f**) Interaction of L20e with ES39L in *S. cerevisiae* (**e**) and *H. sapiens* (**f**).

Single-stranded portions of ES31L and ES39L are central to the two prominent ES clusters found on the surface of the LSU of lower eukaryotes like yeast and *Tetrahymena* (Ben-Shem et al., 2011; Klinge et al., 2011). It has been noted, that non-helical parts provide a platform for the binding of r-proteins and that these interactions contribute to the intertwined nature of the eukaryote-specific layer found in the lower eukaryotic ribosome (Ben-Shem et al., 2011). The non-helical linkers of ES31L and ES39L caused major problems during rRNA modeling in cryo-EM maps above 5 Å (Figure 11b,c). Because of unreliable electron density for these rRNA stretches, the ES31L linker was left out of the cryo-EM based model (Figure 11b) and density for the non-helical parts of ES39L was interpreted in a way, that resulted in wrong connectivity of the strands within the ES (Figure 11c). In contrast to this, better map quality allowed modeling of corresponding parts in the human ribosome (Figure 13b,c). A comparison of the r-protein interacting single-stranded regions of ES31L and ES39L reveals that the stacking interactions formed between specific rRNA

bases and aromatic protein side chains are conserved between yeast and human. Examples include binding of L2 to ES31L via aromatic stacking interactions of two tyrosines (Tyr40/Tyr89 in yeast and human) with U residues of the non-helical rRNA (U2550/U2551 in yeast; U4117/U4118 in human) (Figure 14a,b) or interaction of L13 with ES39L, which involves Tyr167/Tyr168 – A3180/C3181 in *S. cerevisiae* and His167/Tyr168 – C4757/U4758 in *H. sapiens* (Figure 14c,d). Notably, conformations of all rRNA nts involved in these interactions are in excellent agreement with the human 80S electron density (Figure 13b,c). In a third example, L20e is anchored to ES39L via stacking interactions in yeast (Figure 14e). Although human L20e is also bound to ES39L, details of this interaction differ in comparison to *S. cerevisiae* (Figure 14f). Stacking interactions between bases and aromatic protein side chains are common to the mode of mRNA recognition of a diverse set of proteins, including the poly-A binding protein (PABP) (Deo et al., 1999) or the translational regulators sex-lethal (SXL) (Handa et al., 1999) and pumilio (Wang et al., 2002). In an architectural context, non-helical RNA as a platform for protein binding can also be found in small nuclear ribonucleoprotein particles (snRNPs), where Sm proteins assemble onto the single-stranded Sm-site RNA (Pomeranz Krummel et al., 2009). Taken together, stabilization of r-protein and rRNA elements via stacking interactions, which is contributing to the eukaryote-specific entangled layer, is conserved from yeast to human and thus appears to be a general feature of the eukaryotic 80S ribosome.

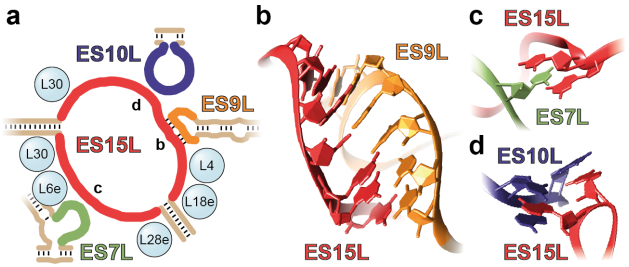


Figure 15 | ES interactions in the human LSU rRNA. (**a**) Schematic view of the ES cluster formed by ES7L, ES9L, ES10L and ES15L. Non-helical rRNA stretches are colored distinctly. (**b**) Pseudoknot between ES9L and ES15L. (**c**,**d**) Interactions of ES7L/ES15L (**c**) and ES10L/ES15L mediated by aromatic base stacking.

In addition to single-stranded regions of ES31L and ES39L, which are used as protein binding sites, the human ribosome contains additional extended non-helical stretches in ES7L, ES9L, ES10L and ES15L. They collectively form an rRNA cluster on the back of the LSU that is not observed in the lower eukaryotic ribosome (Ben-Shem et al., 2011, Klinge et al., 2011). This structural feature has previously been observed in an 8.7 Å cryo-EM reconstruction of the dog 80S ribosome, but due to the limited resolution no molecular model was presented (Chandramouli et al., 2008). The human 80S ribosome structure reveals that the cluster is organized around a drastically enlarged internal loop of ES15L that is very small in *Drosophila* and absent in lower eukaryotes (Anger et al., 2013). This portion of ES15L provides anchor points for several r-proteins (L4, L6e, L18e, L28e, L30) (Figure 15a). Interestingly, the proteins are only found in the periphery of the cluster but do not contribute to the internal stabilization of the structure. In contrast to this, the central region of the cluster is exclusively formed by non-helical rRNA elements of ES7L, ES9L, ES10L and ES15L. One of the ES15L linkers is engaged in a tight interaction with the terminal loop of ES9L via base-pairs and the resulting pseudoknot appears to be a central stabilizing factor for the ES cluster (Figure 15b). This architecture is reminiscent of the ES3S-ES6S interaction in the SSU and appears to be unique for the mammalian ribosome. The remaining single-stranded parts of ES7L and ES10L are anchored to ES15L in an unusual way. Individual bases of the ES non-helical elements are found to be engaged in aromatic stacking interactions with each other (Figure 15c,d), analogous to the interaction of ES31L and ES39L with r-proteins L2, L13 and L14e (Figure 14).

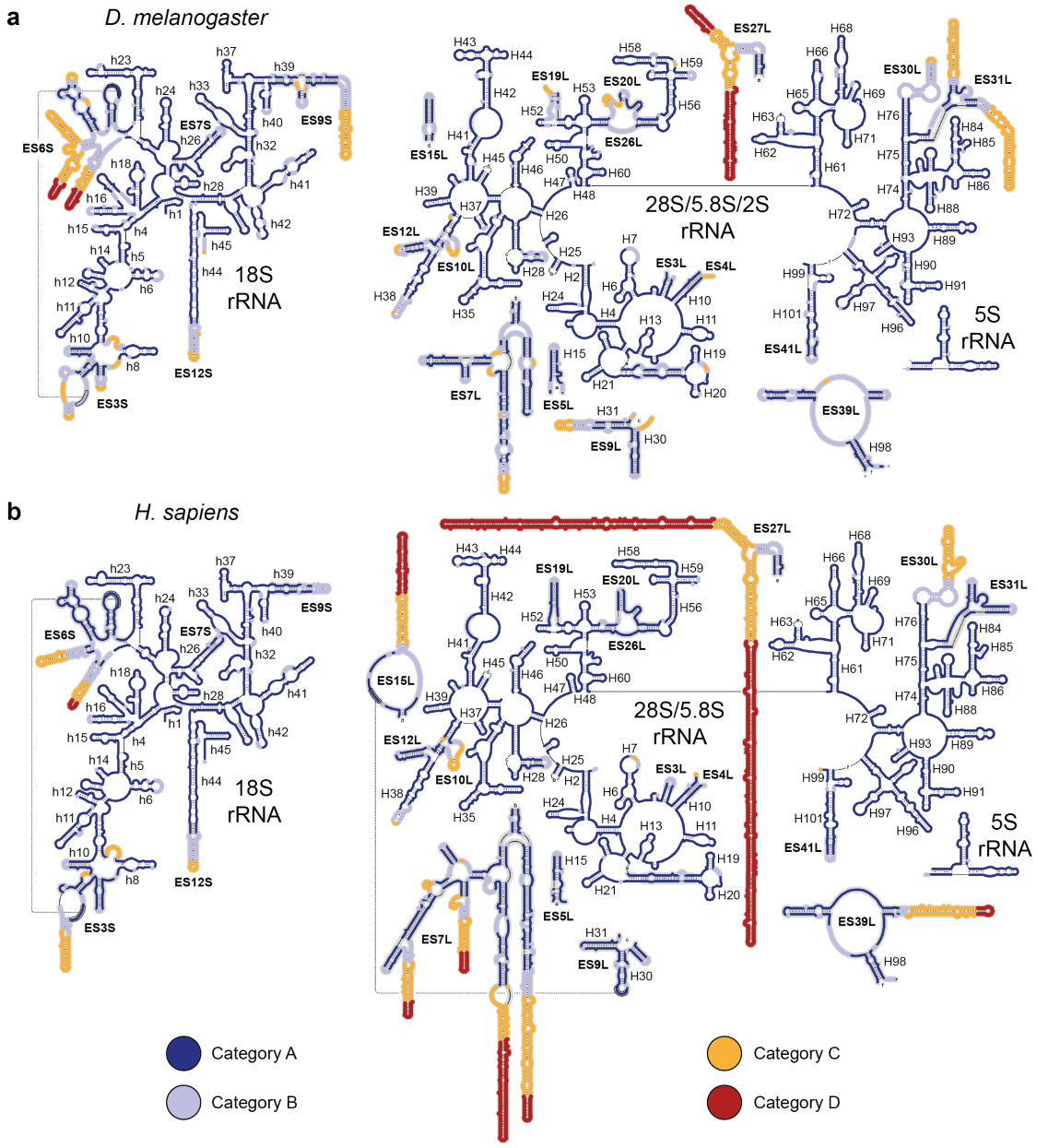


Figure 16 | Reliability of the *Drosophila* and human rRNA models. (**a**), Secondary structure diagram of the *D. melanogaster* rRNA with the reliability colored in four categories (A to D). A (dark blue) represents unambiguously modeled parts of the structure. B (light blue) includes parts of the structure with clear path of the backbone but uncertain base conformations. C (orange) is supported by weak electron density, while D (red) includes parts of the structure that are solely based on secondary structure predictions. (**b**), Secondary structure diagram of the human rRNA with reliability colored as in (a). Categories A to D are saved in PDB entries 3J3C and 3J3E for *D. melanogaster* and 3J3D and 3J3F for *H. sapiens* using the b-factor (BF) values (A: BF=10; B: BF=30; C: BF=70; D: BF=100).

The only mammalian (rabbit) 40S crystal structures available to date are very limited in resolution (7.9 to 9 Å) (Lomakin and Steitz, 2013) and not suitable for a detailed comparison with our higher eukaryote models. In fact, the human 40S structure presented in this work was used as a search template for phase determination of the rabbit 40S electron densities by molecular replacement (Lomakin and Steitz, 2013). Without a high-resolution reference structure in hand, a subjective quality assessment is the only remaining option.

The high degree of conservation between r-proteins from lower and higher eukaryotes in combination with the modest mass increase of r-proteins from higher eukaryotes, results in very reliable r-protein models of the human and *Drosophila* 80S ribosomes, which are almost entirely based on the yeast crystal structure (Ben-Shem et al., 2011). Residual uncertainties remain about side chain orientations and register of some peripheral r-protein portions where resolution becomes limiting but not about location of the proteins *per se*. The situation is more complex for human and *Drosophila* rRNAs. Both include extended parts that have no corresponding (core) structure in the yeast X-ray model and hence required a large amount of *de novo* modeling. To determine the degree of reliability for the resulting rRNA models, nucleotides were systematically grouped in four categories (A to D) (Figure 16, Tables 2 and 3). Category A includes parts of the structures that are in perfect agreement with the electron density. Category B contains stretches with a backbone path that is clearly supported by the map but uncertain orientations of the bases. Category C is supported by electron density that lacks detail information but still allows tracing of the principal rRNA path. All remaining nucleotides, which lie outside of the map and are modeled solely based on secondary structure predictions, are grouped into category D. The vast majority of the rRNA models fall into the two best categories A and B, with 80 and 90% for the human and *Drosophila* structures, respectively (Tables 2 and 3). Importantly, all single-stranded linkers and branch points of *de novo* modeled ES are included in this fraction. The remaining weaker parts of the structures in categories C and D appear to be simple unbranched helices in most cases (Figure 16). This conclusion is based on the observation of (weak) electron density supporting model parts in category C as well as focused secondary structure predictions for regions C and D. Taken together, the human and *Drosophila* rRNA models presented in this work are very likely to be correct, with uncertainties only remaining for base orientations in some parts of the structures.

Table 2 | Reliability of the *D. melanogaster* rRNA models.

		Category A			Category B			Category C			Category D	
rRNA	length (nts)	nts	%	%(c)	nts	%	%(c)	nts	%	%(c)	nts	%
18S	1995	1464	73	75	298	15	15	195	10	10	38	2
5S	120	117	98	98	3	3	3	0	0	0	0	0
5.8S	123	113	92	92	10	8	8	0	0	0	0	0
2S	30	29	97	97	1	3	3	0	0	0	0	0
28S	3925	2980	76	78	537	14	14	289	7	8	119	3
Total	6193	4703	76	78	849	14	14	484	8	8	157	3

In addition to the percent values for the categories based on the total number of nucleotides, a corrected (c) percent value is given for categories A-C. This is calculated with the total number of nucleotides corrected by nucleotides of category D, since they are not visible in the structure.

Table 3 | Reliability of the *H. sapiens* rRNA models.

		Category A			Category B			Category C			Category D	
rRNA	length (nts)	nts	%	%(c)	nts	%	%(c)	nts	%	%(c)	nts	%
18S	1869	1581	85	85	176	9	9	104	6	6	8	0
5S	121	118	98	98	3	2	2	0	0	0	0	0
5.8S	157	140	89	89	14	9	9	3	2	2	0	0
28S	5070	3391	67	79	366	7	9	541	11	13	772	15
Total	7217	5230	72	81	559	8	9	648	9	10	780	11

In addition to the percent values for the categories based on the total number of nucleotides, a corrected (c) percent value is given for categories A-C. This is calculated with the total number of nucleotides corrected by nucleotides of category D, since they are not visible in the structure.

4.3 Structural Evolution of the Ribosome

The ribosomal core. The discovery that RNA can not only carry genetic information but is also capable to function as a catalyst (Cech et al., 1981; Guerrier-Takada et al., 1983) led to the so-called RNA world hypothesis (Gilbert, 1986). Today it is widely accepted that the ribosome and the process of translation originated from this ancient RNA world that predates splitting of the three domains of life (Noller, 2012). As a consequence, all modern ribosomes contain a common rRNA core that performs the two fundamental functions: decoding and peptidyl transfer. The PTC within domain V of the LSU rRNA is thought to constitute the oldest part of the ribosome and later additions during evolution are located towards the surface of the particle (Fox and Ashinikumar, 2004). In line with this, comparative analysis suggests that the LSU core is built up in concentric shells with the PTC as origin (Hsiao et al., 2009). The PTC itself is formed by a symmetrical rRNA piece that binds the 3'-CCA ends of A- and P-tRNAs (Nissen et al., 2000). Symmetry and similarities between the two halves of the PTC indicate that the ribosomal core and the A- and P-tRNA binding sites initially arose from a duplication event of the same RNA fragment (Agmon et al., 2003). The E-site, on the other hand, is thought to be a later addition to the ribosome (Steitz, 2008). This is in agreement with studies that attempt to give insights into historical timing of LSU evolution based on interconnectivity of structural elements. Older parts had more time to be integrated in the structure during evolution and thus are predicted to be more interconnected than younger portions. Results show that LSU rRNA domain V together with parts of domains II and IV are highly interconnected, while domains I and III appear to be later additions (Hury et al., 2006). The oldest regions largely overlap with the minimal ribosome parts identified by comparative analysis (Mears et al., 2002). More recently A-minor interactions were analyzed to obtain a hierarchical model of the LSU rRNA evolution (Bokov and Steinberg, 2009). A two-component interaction like the A-minor motif can be seen as a timing event if one component predates the other. In concrete terms, conformational integrity of A-minor motifs is dependent on the presence of the acceptor helices prior to the emergence of flipped out adenines. The study confirms the old age of domain V (including the PTC) forming the ribosomal inner core. The core was later extended by parts of domain II, followed by domain IV. It also allows the conclusion that the L1 stalk and GTPase center (translation factor binding site) were late additions to the ribosome (Bokov and Steinberg, 2009). Late invention of a translation factor system during evolution is in agreement with the idea that the ribosome essentially functions as a brownian motor for tRNA movement and that EF-G is ancillary rather than causative in promoting translocation (Frank and Gonzalez, 2010; Frank, 2012). LSU rRNA domain IV forms major contacts with the SSU and the emergence of this domain during evolution likely corresponds to the beginning of the SSU. The SSU might not have originated in later times as an addition to the growing ribosome but rather evolved separately as a replicase in the RNA world. After being added to the ribosome such a replicase could function in moving of the translation template. In such a scenario, the SSU decoding site would be a relic of an RNA replication mechanism, which used A-minor interactions to monitor accuracy of replication (Noller, 2012). An important question is, what drove evolution of translation to produce polypeptides from an RNA world. An interesting hypothesis is based on observations that binding of short peptides to RNA can cause large-scale structural changes. The presence of peptides could provide a substantial advantage by allowing more diverse RNA folds and thus expanded functional capabilities. In this line of argumentation translation would have initially evolved not to create a protein world but to allow a more diverse RNA world (Noller, 2012). Collectively, RNA and protein coevolution resulted in the universally conserved core of all modern cytosolic ribosomes, which is build up from 34 r-proteins and around 4400 nts in 3 rRNA molecules (Melnikov et al., 2012).

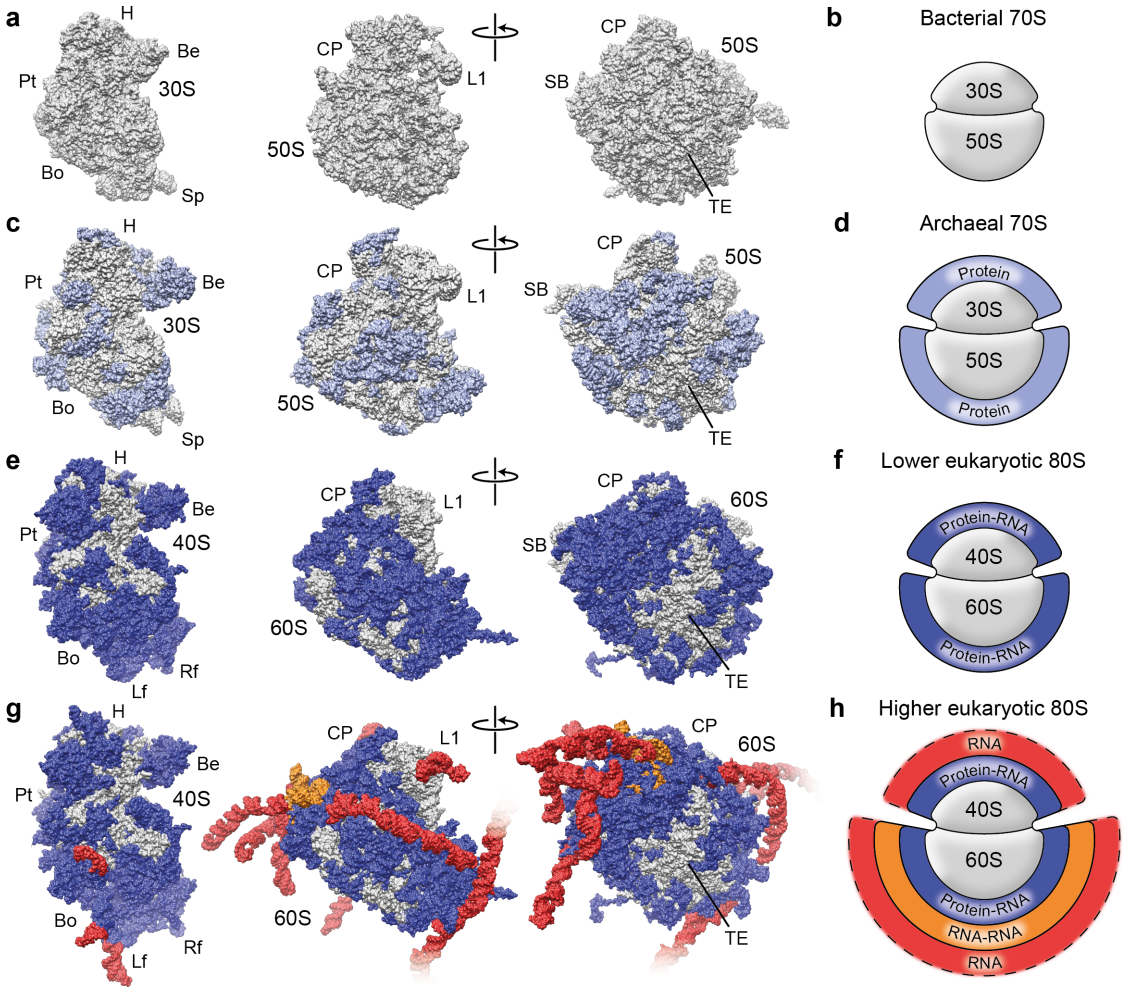


Figure 17 | Structural evolution of the ribosome. (**a-h**), Surface representations (**a**,**c**,**e**,**g**) and schematics (**b**,**d**,**f**,**h**) of the bacterial *T. thermophilus* 70S ribosome (**a**,**b**) (Jenner et al., 2010), the *P. furiosus* 70S ribosome (**c**,**d**) (Armache et al., 2013) (the archaea-specific layer is shown), the *S. cerevisiae* 80S ribosome (**e**,**f**) (Ben-Shem et al., 2011) (the eukaryote-specific protein-RNA layer is shown), and the mammalian 80S ribosome from *H. sapiens* (**g**,**h**) (Anger et al., 2013) (the two additional layers RNA-RNA and RNA-only, are shown). Note that bacteria-specific proteins have not been colored separately in (a) for clarity.

Beyond the ribosomal core. Specific features on both rRNA and r-protein sides further extend the ribosomal core in each domains of life. Although bacteria contain 23 domain-specific r-proteins (S1, S6, S16, S18, S20, S21, S22, THX, L9, L12/7, L17, L19, L20, L21, L25, L27, L28, L31, L32, L33, L34, L35 and L36), the majority of their ribosome structure is conserved in eukaryotes and thus can be considered to form the core of the 80S ribosome (Figure 17a,b). The few examples of bacteria-specific rRNA features include the unique architecture of h33, as well as conformations of h16, H15, H30, H58 and the L1-stalk. Archaeal ribosomes are of intermediate complexity compared to bacteria and eukaryotes. Here, extension of the ribosome core is largely due to 34 r-proteins (S1e, S4e, S6e, S8e, S17e, S19e, S24e, S25e, S26e, S27e, S28e, S30e, S31e, L13e, L14e, L15e, L18e, L19e, L21e, L24e, L30e, L31e, L32e, L33e, L34e, L37e, L38e, L39e, L40e, L41e, L43e, L44e, P1 and P2) that are also shared with eukaryotes (Figure 17c,d). There is only a limited number of known archaea-specific r-proteins such as LX (Greber et al., 2012a) or three LSU proteins (L45a, L46a and L47a) that have been identified by proteomic characterization of ribosomal subunits (Márquez et al., 2011). On the rRNA side, archaea contain a number of very

small VR and ES that show the same basic architecture as the corresponding parts in eukaryotic rRNA (Armache et al., 2013; Ban et al., 2000; Greber et al., 2012a). Archaeal rRNA can thus be regarded as a chimera between bacteria and eukaryotes. Notably, the changed geometry of several rRNA parts in comparison to bacteria involves stabilization by promiscuous archaeal r-proteins identified in this work. For instance, L8e and L8e(S) bind to remodeled H15 and h33, respectively and the changed rRNA conformation around H54/H58 is stabilized by L14e(2). Additional 12 r-proteins (S7e, S10e, S12e, S21e, RACK1, L6e, L20e, L22e, L27e, L28e, L29e and L36e) are found to be eukaryote-specific. S12e and L27e occupy positions of the related archaeal proteins L8e(S) and L14e(2), suggesting that these eukaryotic r-proteins evolved through increased copy number and binding site promiscuity. Structures of lower eukaryotic ribosomes have revealed that the additional eukaryotic-specific r-proteins and r-protein extensions form a network of interactions with the rRNA ES, resulting in an intertwined RNA-protein layer (Figure 17e,f). In higher eukaryotes, this RNA-protein layer has increased in size and complexity owing to the presence of additional r-protein extensions and rRNA ES insertions (Figure 17g,h). Moreover, the substantial increase in RNA mass of higher eukaryotes, particularly mammalian ribosomes, compared to lower eukaryotes, has resulted in the presence of two additional RNA layers (Figure 17g,h). A rigid inner layer results from multiple RNA-RNA tertiary interactions such as the aromatic stacking interactions of ES7L/ES10L/ES15L as well as the pseudoknot formed by ES9L/ES15L. A second flexible outer layer arises from helical insertions and extensions of the rRNA ES. The striking observation of two additional RNA layers in higher eukaryotic 80S ribosomes raises more questions than it answers. It will be very interesting to elucidate the benefits of an RNA-only layer from both a structural and functional perspective. Why did r-proteins not coevolve together with the rRNA beyond a certain point in the higher eukaryotic ribosome and what consequences does the presence of very long RNA tentacles have on various aspects of ribosome biology? For instance, the diffusion coefficient of *E. coli* ribosomes has been determined to be 0.04 $\mu m^2 s^{-1}$ (Bakshi et al., 2012). Given the larger size of eukaryotic ribosomes and the inverse proportionality of diffusion coefficient and hydrodynamic radius this value can be expected to be lower for 80S ribosomes with the most dramatic effect for mammalians due to presence of the rRNA tentacles. Other interesting topics include membrane association of mammalian 80S ribosomes during protein translocation across the endoplasmatic reticulum membrane and their organization in polysomes. To what extent do rRNA tentacles influence these processes? Initial structural studies using cryoelectron tomography could not give an answer to this question because high flexibility of the extended rRNA parts prevented their visualization (Brandt et al., 2010; Pfeffer et al., 2012). The long rRNA tentacles of higher eukaryotic ribosomes could be decorated with numerous factors, which might serve to protect the structures from degradation by cellular RNases and could harbor a functional repertoire specific for the mammalian ribosome. Taken together, mining of the higher eukaryotic ribosome models presented in this dissertation for biological function will be an important challenge for future research, especially since the function of ES remains one of the major open questions in the ribosome field.

5 Appendix

This section includes tables summarizing the recently revised nomenclature for r-proteins (Ben-Shem et al., 2011; Jenner et al., 2012) (Supplementary Tables S1 and S2). In contrast to the original proposal eukaryotic r-protein P0 is named L10, as suggested by Liljas, Moore and Yusupov (www.elsevierblogs.com/currentcomments/?p=686). In addition, a complete list for r-protein PDB chain IDs is included to facilitate comparison of the currently available eukaryotic ribosome structures (Supplementary Tables S3 and S4).

Table S1 | Small ribosomal subunit r-protein nomenclature and taxonomic distribution.

Protein name	Taxonomic range			Bacteria name	Old yeast name	Old human name
S1	B*			S1	(-)	(-)
S1e		A	E	(-)	S1	S3A
S2	B	A	E	S2	S0	SA
S3	B	A	E	S3	S3	S3
S4	B	A	E	S4	S9	S9
S4e		A	E	(-)	S4	S4
S5	B	A	E	S5	S2	S2
S6	B			S6	(-)	(-)
S6e		A	E	(-)	S6	S6
S7	B	A	E	S7	S5	S5
S7e			E	(-)	S7	S7
S8	B	A	E	S8	S22	S15A
S8e		A	E	(-)	S8	S8
S9	B	A	E	S9	S16	S16
S10	B	A	E	S10	S20	S20
S10e			E	(-)	S10	S10
S11	B	A	E	S11	S14	S14
S12	B	A	E	S12	S23	S23
S12e			E	(-)	S12	S12
S13	B	A	E	S13	S18	S18
S14	B	A	E	S14	S29	S29
S15	B	A	E	S15	S13	S13
S16	B			S16	(-)	(-)
S17	B	A	E	S17	S11	S11
S17e		A	E	(-)	S17	S17
S18	B			S18	(-)	(-)
S19	B	A	E	S19	S15	S15
S19e		A	E	(-)	S19	S19
S20	B			S20	(-)	(-)
S21	B*			S21	(-)	(-)
S22	B*			S22	(-)	(-)
S21e			E*	(-)	S21	S21
S24e		A	E	(-)	S24	S24
S25e		A*	E	(-)	S25	S25
S26e		A*	E	(-)	S26	S26
S27e		A	E	(-)	S27	S27
S28e		A*	E	(-)	S28	S28
S30e		A*	E	(-)	S30	S30
S31e		A	E	(-)	S31	S27A
RACK1			E	(-)	Asc1	RACK1
THX	B*			THX	(-)	(-)

An asterisk indicates the presence of a protein family in some, but not all, representatives of a domain. Abbreviations: A, archaea; B, bacteria; E, eukaryotes.

Table S2 | Large ribosomal subunit r-protein nomenclature and taxonomic distribution.

Protein name	Taxonomic range			Bacteria name	Old yeast name	Old human name
L1	B	A	E	L1	L1	L10A
L2	B	A	E	L2	L2	L2
L3	B	A	E	L3	L3	L3
L4	B	A	E	L4	L4	L4
L5	B	A	E	L5	L11	L11
L6	B	A	E	L6	L9	L9
L6e			E	(-)	L6	L6
L8e	B*	A	E	(-)	L8	L7A
L9	B			L9	(-)	(-)
L10	B	A	E	L10	P0	P0
L11	B	A	E	L11	L12	L12
L12/L7	B			L12/L7	(-)	(-)
L13	B	A	E	L13	L16	L13A
L13e		A*	E	(-)	L13	L13
L14	B	A	E	L14	L23	L23
L14e		A*	E*	(-)	L14	L14
L15	B	A	E	L15	L28	L27A
L15e		A	E	(-)	L15	L15
L16	B	A	E	L16	L10	L10
L17	B			L17	(-)	(-)
L18	B	A	E	L18	L5	L5
L18e		A	E	(-)	L18	L18
L19	B			L19	(-)	(-)
L19e		A	E	(-)	L19	L19
L20	B			L20	(-)	(-)
L20e			E	(-)	L20	L18A
L21	B			L21	(-)	(-)
L21e		A	E	(-)	L21	L21
L22	B	A	E	L22	L17	L17
L22e			E	(-)	L22	L22
L23	B	A	E	L23	L25	L23A
L24	B	A	E	L24	L26	L26
L24e		A	E	(-)	L24	L24
L25	B*			L25	(-)	(-)
L27	B			L27	(-)	(-)
L27e			E	(-)	L27	L27
L28	B			L28	(-)	(-)
L28e			E*	(-)	(-)	L28
L29	B	A	E	L29	L35	L35
L29e			E	(-)	L29	L29
L30	B*	A	E	L30	L7	L7
L30e		A*	E	(-)	L30	L30
L31	B			L31	(-)	(-)
L31e		A	E	(-)	L31	L31
L32	B			L32	(-)	(-)
L32e		A	E	(-)	L32	L32
L33	B			L33	(-)	(-)
L33e		A*	E	(-)	L33	L35A
L34	B			L34	(-)	(-)
L34e		A*	E	(-)	L34	L34
L35	B			L35	(-)	(-)
L36	B			L36	(-)	(-)
L36e			E	(-)	L36	L36
L37e		A*	E	(-)	L37	L37
L38e		A*	E	(-)	L38	L38
L39e		A	E	(-)	L39	L39
L40e		A*	E	(-)	L40	L40

Table S2 | (continued).

Protein name	Taxonomic range		Bacteria name	Old yeast name	Old human name
L41e	A*	E	(-)	L41	L41
L43e	A	E	(-)	L43	L37A
L44e	A	E	(-)	L42	L36A
P1	A	E	(-)	P1 (αβ)	LP1
P2	A	E	(-)	P2 (αβ)	LP2
LX	A*		(-)	(-)	(-)

An asterisk indicates the presence of a protein family in some, but not all, representatives of a domain. Abbreviations: A, archaea; B, bacteria; E, eukaryotes.

Table S3 | Small ribosomal subunit r-protein PDB chain IDs.

	Archaea	Eukarya						
Protein name	*Pfu* 3J20.pdb	*Tbr* 3ZEY.pdb	*Tth* 2XZM/N.pdb	*Sce* 3U5C/G.pdb	*Sce* 3IZB.pdb	*Tae* 3IZ6.pdb	*Dme* 3J38.pdb	*Hsa* 3J3A.pdb
S1e	A	0	4	B	(-)	(-)	B	B
S2	B	C	B	A	A	A	A	A
S3	C	X	C	D	B	B	D	D
S4	D	6	D	J	C	C	J	J
S4e	E	1	W	E	D	D	E	E
S5	F	P	E	C	E	E	C	C
S6e	G	3	Y	G	(-)	(-)	G	G
S7	H	2	G	F	F	F	F	F
S7e	(-)	4	3	H	G	G	H	H
S8	I	J	H	W	H	H	W	W
S8e	J	5	2	I	(-)	(-)	I	I
S9	K	K	I	Q	I	I	Q	Q
S10	L	Q	J	U	J	J	U	U
S10e	(-)	D	7	K	(-)	(-)	K	K
S11	M	H	K	O	K	K	O	O
S12	N	S	L	X	L	L	X	X
S12e	(-)	F	U	M	(-)	(-)	M	M
S13	O	M	M	S	M	M	S	S
S14	P	8	N	d	N	N	d	d
S15	Q	G	O	N	O	O	N	N
S17	R	E	Q	L	P	P	L	L
S17e	S	L	V	R	Q	Q	R	R
S19	T	I	S	P	R	R	P	P
S19e	U	O	T	T	S	S	T	T
S21e	(-)	R	Z	V	T	T	V	V
S24e	V	T	P	Y	U	U	Y	Y
S25e	(-)	U	8	Z	V	V	Z	Z
S26e	(-)	V	5	a	W	W	a	a
S27e	W	W	6	b	X	X	b	b
S28e	X	Z	1	c	Y	Y	c	c
S30e	(-)	Y	X	e	Z	Z	e	e
S31e	Y	9	9	f	(-)	(-)	f	f
RACK1	(-)	7	R	g	a	a	g	g
L8eS	3	(-)	(-)	(-)	(-)	(-)	(-)	(-)

Abbreviations: *Dme*, *Drosophila melanogaster*; *Hsa*, *Homo sapiens*; *Pfu*, *Pyrococcus furiosus*; *Sce*, *Saccharomyces cerevisiae*; *Tae*, *Tritcum aestivum*; *Tbr*, *Trypanosoma brucei*; *Tth*, *Tetrahymena thermophila*. References: *Dme* (3J38.pdb) (Anger et al., 2013), *Hsa* (3J3A.pdb) (Anger et al., 2013), *Pfu* (3J20.pdb) (Armache et al., 2013), *Sce* (3IZB.pdb) (Armache et al., 2010b), *Sce* (3U5C/G.pdb) (Ben-Shem et al., 2011), *Tae* (3IZ6.pdb) (Armache et al., 2010b), *Tbr* (3ZEY.pdb) (Hashem et al., 2013a), *Tth* (2XZM/N.pdb) (Rabl et al., 2011).

Table S4 | Large ribosomal subunit r-protein PDB chain IDs.

	Archaea	Eukarya						
Protein name	*Pfu* 3J21.pdb	*Tbr* 3ZF7.pdb	*Tth* 4A1A/B/C/D/E/7/8/9.pdb	*Sce* 3U5E/I.pdb	*Sce* 3IZS.pdb	*Tae* 3IZR.pdb	*Dme* 3J39.pdb	*Hsa* 3J3B.pdb
L1	A	J	(-)	(-)	A	A	z	z
L2	B	e	4A1A/C/E/7_A	A	B	B	A	A
L3	C	f	4A1A/C/E/7_B	B	C	C	B	B
L4	D	r	4A1A/C/E/7_C	C	D	D	C	C
L5	E	L	4A1A/C/E/7_D	J	E	E	J	J
L6	F	y	4A1A/C/E/7_E	H	F	F	H	H
L6e	(-)	v	4A1B/D/8/9_E	E	G	G	E	E
L8e	G	x	4A1A/C/E/7_F	G	H	H	G	G
L10	k	(-)	4A1A/C/E/7_G	3U5I_q	s	s	q	q
L11	H	M	(-)	3U5I_K	J	J	K	K
L13	I	O	4A1A/C/E/7_I	O	K	K	O	O
L13e	(-)	N	4A1B/D/8/9_U	L	L	L	L	L
L14	J	W	4A1A/C/E/7_J	V	M	M	V	V
L14e	K	P	4A1B/D/8/9_F	M	N	N	M	M
L15	L	b	4A1A/C/E/7_K	a	O	O	a	a
L15e	M	Q	4A1A/C/E/7_L	N	P	P	N	N
L16	N	K	4A1A/C/E/7_H	I	I	I	I	I
L18	O	u	4A1A/C/E/7_M	D	Q	Q	D	D
L18e	P	I	4A1A/C/E/7_N	Q	R	R	Q	Q
L19e	Q	T	4A1A/C/E/7_O	R	T	T	R	R
L20e	(-)	S	4A1B/D/8/9_X	S	S	S	S	S
L21e	R	U	4A1A/C/E/7_P	T	U	U	T	T
L22	S	R	4A1A/C/E/7_Q	P	V	V	P	P
L22e	(-)	V	4A1B/D/8/9_M	U	W	W	U	U
L23	T	X	4A1A/C/E/7_R	X	X	X	X	X
L24	U	Z	4A1A/C/E/7_S	Y	Y	Y	Y	Y
L24e	V	Y	4A1A/C/E/7_T	W	Z	Z	W	W
L27e	(-)	a	4A1B/D/8/9_N	Z	a	a	Z	Z
L28e	(-)	c	4A1B/D/8/9_O	(-)	(-)	b	r	r
L29	W	k	4A1A/C/E/7_U	h	c	c	h	h
L29e	(-)	d	4A1B/D/8/9_T	b	d	d	b	b
L30	Y	w	4A1A/C/E/7_V	F	e	e	F	F
L30e	Z	g	4A1B/D/8/9_G	c	f	f	c	c
L31e	a	h	4A1A/C/E/7_W	d	g	g	d	d
L32e	b	i	4A1A/C/E/7_X	e	h	h	e	e
L33e	c	l	4A1B/D/8/9_H	f	j	j	f	f
L34e	d	j	4A1B/D/8/9_L	g	i	i	g	g
L36e	(-)	m	4A1B/D/8/9_Q	i	k	k	i	i
L37e	e	n	4A1B/D/8/9_A	j	l	l	j	j
L38e	(-)	p	4A1B/D/8/9_P	k	n	n	k	k
L39e	f	q	4A1B/D/8/9_B	l	o	o	l	l
L40e	g	s	4A1B/D/8/9_K	m	p	p	m	m
L41e	h	(-)	(-)	n	q	q	n	n
L44e	j	t	4A1B/D/8/9_C	o	r	r	o	o
L43e	i	o	4A1A/C/E/7_Y	p	m	m	p	p
P1	(-)	(-)	(-)	3U5I_r	t / u	t / u	(-)	s / t
P2	(-)	(-)	(-)	3U5I_s	v / w	v / w	(-)	u / v
L8e2	4	(-)	(-)	(-)	(-)	(-)	(-)	(-)
L14e2	5	(-)	(-)	(-)	(-)	(-)	(-)	(-)
LX	l	(-)	(-)	(-)	(-)	(-)	(-)	(-)
S24eL	6	(-)	(-)	(-)	(-)	(-)	(-)	(-)

Abbreviations: *Dme*, *Drosophila melanogaster*; *Hsa*, *Homo sapiens*; *Pfu*, *Pyrococcus furiosus*; *Sce*, *Saccharomyces cerevisiae*; *Tae*, *Tritcum aestivum*; *Tbr*, *Trypanosoma brucei*; *Tth*, *Tetrahymena thermophila*. References: *Dme* (3J39.pdb) (Anger et al., 2013), *Hsa* (3J3B.pdb) (Anger et al., 2013), *Pfu* (3J21.pdb) (Armache et al., 2013), *Sce* (3IZS.pdb) (Armache et al., 2010b), *Sce* (3U5E/I.pdb) (Ben-Shem et al., 2011), *Tae* (3IZR.pdb) (Armache et al., 2010b), *Tbr* (3ZF7.pdb) (Hashem et al., 2013a), *Tth* (4A1A/B/C/D/E/7/8/9.pdb) (Klinge et al., 2011).

6 References

Agmon, I. et al. (2003). On peptide bond formation, translocation, nascent protein progression and the regulatory properties of ribosomes. Eur. J. Biochem. *270*, 2543–2556.

Agrawal, R.K. et al. (1998). Visualization of elongation factor G on the Escherichia coli 70S ribosome: the mechanism of translocation. Proc. Natl. Acad. Sci. U.S.A. *95*, 6134–6138.

Alkalaeva, E.Z. et al. (2006). In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. Cell *125*, 1125–1136.

Alkemar, G., and Nygård, O. (2003). A possible tertiary rRNA interaction between expansion segments ES3 and ES6 in eukaryotic 40S ribosomal subunits. RNA *9*, 20–24.

Alkemar, G., and Nygård, O. (2006). Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. Biochemistry *45*, 8067–8078.

Andersen, C.B.F. et al. (2006). Structure of eEF3 and the mechanism of transfer RNA release from the E-site. Nature *443*, 663–668.

Anger, A.M. et al. (2013). Structures of the human and *Drosophila* 80S ribosome. Nature *497*, 80–85.

Armache, J.-P. et al. (2010a). Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. Proc. Natl. Acad. Sci. U.S.A. *107*, 19748–19753.

Armache, J.-P. et al. (2010b). Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. Proc. Natl. Acad. Sci. U.S.A. *107*, 19754–19759.

Armache, J.-P. et al. (2013). Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. Nucleic Acids Res. *41*, 1284–1293.

Bai, X.-C. et al. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. Elife *2*, e00461.

Bakshi, S. et al. (2012). Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Mol. Microbiol. *85*, 21–38.

Ban, N. et al. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science *289*, 905–920.

Becker, T. et al. (2009). Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. Science *326*, 1369–1373.

Becker, T. et al. (2011). Structure of the no-go mRNA decay complex Dom34–Hbs1 bound to a stalled 80S ribosome. Nat. Struct. Mol. Biol. *18*, 715–720.

Becker, T. et al. (2012). Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. Nature *482*, 501–506.

Beckmann, R. et al. (2001). Architecture of the protein-conducting channel associated with the translating 80S ribosome. Cell *107*, 361–372.

Benelli, D., and Londei, P. (2011). Translation initiation in Archaea: conserved and domain-specific features. Biochem. Soc. Trans. *39*, 89–93.

Ben-Shem, A. et al. (2010). Crystal structure of the eukaryotic ribosome. Science *330*, 1203–1209.

Ben-Shem, A. et al. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. Science *334*, 1524–1529.

Blau, M. et al. (2005). ERj1p uses a universal ribosomal adaptor site to coordinate the 80S ribosome at the membrane. Nat. Struct. Mol. Biol. *12*, 1015–1016.

Bokov, K., and Steinberg, S.V. (2009). A hierarchical model for evolution of 23S ribosomal RNA. Nature *457*, 977–980.

Bommer, U.A. et al. (1991). Eukaryotic initiation factors eIF-2 and eIF-3: interactions, structure and localization in ribosomal initiation complexes. Biochimie *73*, 1007–1019.

Bradatsch, B. et al. (2012). Structure of the pre-60S ribosomal subunit with nuclear export factor Arx1 bound at the exit tunnel. Nat. Struct. Mol. Biol. *19*, 1234–1241.

Brandt, F. et al. (2009). The native 3D organization of bacterial polysomes. Cell *136*, 261–271.

Brandt, F. et al. (2010). The three-dimensional organization of polyribosomes in intact human cells. Mol. Cell *39*, 560–569.

Brodersen, D.E. et al. (2002). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. J. Mol. Biol. *316*, 725–768.

Budkevich, T. et al. (2011). Structure and dynamics of the mammalian ribosomal pretranslocation complex. Mol. Cell *44*, 214–224.

Bulygin, K. et al. (2005). The first position of a codon placed in the A site of the human 80S ribosome contacts nucleotide C1696 of the 18S rRNA as well as proteins S2, S3, S3a, S30, and S15. Biochemistry *44*, 2153–2162.

Cannone, J.J. et al. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics *3*, 2.

Cech, T.R. et al. (1981). In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell *27*, 487–496.

Cech, T.R. (2000). Structural biology. The ribosome is a ribozyme. Science *289*, 878–879.

Ceci, M. et al. (2003). Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. Nature *426*, 579–584.

Chandramouli, P. et al. (2008). Structure of the mammalian 80S ribosome at 8.7 Å resolution. Structure *16*, 535–548.

Choli, T. et al. (1993). Isolation and characterization of a new ribosomal protein from the thermophilic eubacteria, *Thermus thermophilus*, *T. aquaticus* and *T. flavus*. Biol. Chem. Hoppe-Seyler *374*, 377–383.

Clark, C.G. (1987). On the evolution of ribosomal RNA. J. Mol. Evol. *25*, 343–350.

Connell, S.R. et al. (2007). Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. Mol. Cell *25*, 751–764.

Connolly, K., and Culver, G. (2009). Deconstructing ribosome construction. Trends Biochem. Sci. *34*, 256–263.

Coyle, S.M. et al. (2009). Direct link between RACK1 function and localization at the ribosome in vivo. Mol. Cell. Biol. *29*, 1626–1634.

Crick, F. et al. (1961). General nature of the genetic code for proteins. Nature *192*, 1227–1232.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563.

DeLabre, M.L.et al. (2002). RPL29 codes for a non-essential protein of the 60S ribosomal subunit in *Saccharomyces cerevisiae* and exhibits synthetic lethality with mutations in genes for proteins required for subunit coupling. Biochim. Biophys. Acta *1574*, 255–261.

Demeshkina, N. et al. (2012). A new understanding of the decoding principle on the ribosome. Nature *484*, 256–259.

Deo, R.C. et al. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835–845.

Desmond, E. et al. (2011). On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. Res. Microbiol. *162*, 53–70.

Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harb. Perspect. Biol. *4*, a013706.

Dinman, J.D. (2009). The eukaryotic ribosome: current status and challenges. J. Biol. Chem. *284*, 11761–11765.

Doerfel, L.K. et al. (2013). EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science *339*, 85–88.

Dube, P. et al. (1998a). Correlation of the expansion segments in mammalian rRNA with the fine structure of the 80 S ribosome; a cryoelectron microscopic reconstruction of the rabbit reticulocyte ribosome at 21 Å resolution. J. Mol. Biol. *279*, 403–421.

Dube, P. et al. (1998b). The 80S rat liver ribosome at 25 Å resolution by electron cryomicroscopy and angular reconstitution. Structure *6*, 389–399.

Dunkle, J.A. et al. (2011). Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. Science *332*, 981–984.

Eickbush, T.H., and Eickbush, D.G. (2007). Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. Genetics *175*, 477–485.

Finley, D. et al. (1989). The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. Nature *338*, 394–401.

Fischer, N. et al. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature *466*, 329–333.

Fox, G.E. (2010). Origin and Evolution of the Ribosome. Cold Spring Harb. Perspect. Biol. *2*, a003483.

Fox, G.E., Ashinikumar, K.N. (2004). The evolutionary history of the translation machinery. In: de Pouplana LR, editor. The genetic code and the origin of life. New York: Kluwer Academic / Plenum. p. 92–105.

Franckenberg, S. et al. (2012). Structural view on recycling of archaeal and eukaryotic ribosomes after canonical termination and ribosome rescue. Curr. Opin. Struct. Biol. *22*, 786–796.

Frank, J. (2012). Intermediate states during mRNA-tRNA translocation. Curr. Opin. Struct. Biol. *22*, 778–785.

Frank, J., and Agrawal, R.K. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature *406*, 318–322.

Frank, J., and Gonzalez, R.L. (2010). Structure and dynamics of a processive Brownian motor: the translating ribosome. Annu. Rev. Biochem. *79*, 381–412.

Freistroffer, D.V. et al. (1997). Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. EMBO J. *16*, 4126–4133.

Fromont-Racine, M. et al. (2003). Ribosome assembly in eukaryotes. Gene *313*, 17–42.

Gao, H. et al. (2005). The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. Proc. Natl. Acad. Sci. U.S.A. *102*, 10206–10211.

Gao, Y.-G. et al. (2009). The structure of the ribosome with elongation factor G trapped in the posttranslocational state. Science *326*, 694–699.

Gartmann, M. et al. (2010). Mechanism of eIF6-mediated inhibition of ribosomal subunit joining. J. Biol. Chem. *285*, 14848–14851.

Gerbi, S.A. (1996). Expansion Segments: Regions of Variable Size that Interrupt the Universal Core Secondary Structure of Ribosomal RNA. CRC Press 71–87.

Gilbert, W. (1986). The RNA world. Nature *319*, 618–618.

Greber, B.J. et al. (2012a). Cryo-EM Structure of the Archaeal 50S Ribosomal Subunit in Complex with Initiation Factor 6 and Implications for Ribosome Evolution. J. Mol. Biol. *418*, 145-160.

Greber, B.J. et al. (2012b). Cryo-EM structures of Arx1 and maturation factors Rei1 and Jjj1 bound to the 60S ribosomal subunit. Nat. Struct. Mol. Biol. *19*, 1228–1233.

Gross, B. et al. (1983). Spatial arrangement of proteins within the small subunit of rat liver ribosomes studied by cross-linking. EMBO J. *2*, 255–260.

Grosso, S. et al. (2008a). PKCbetaII modulates translation independently from mTOR and through RACK1. Biochem. J. *415*, 77–85.

Grosso, S. et al. (2008b). Eukaryotic ribosomes host PKC activity. Biochem. Biophys. Res. Commun. *376*, 65–69.

Guerrier-Takada, C. et al. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849–857.

Gutierrez, E. et al. (2013). eIF5A Promotes Translation of Polyproline Motifs. Mol. Cell *51*, 35–45.

Halic, M. et al. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature *427*, 808–814.

Halic, M. et al. (2005). Localization and dynamic behavior of ribosomal protein L30e. Nat. Struct. Mol. Biol. *12*, 467–468.

Halic, M. et al. (2006). Signal recognition particle receptor exposes the ribosomal translocon binding site. Science *312*, 745–747.

Hancock, J.M. et al. (1988). Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of Drosophila melanogaster. Mol. Biol. Evol. *5*, 393–414.

Handa, N. et al. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. Nature *398*, 579–585.

Harms, J. et al. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell *107*, 679–688.

Hashem, Y. et al. (2013a). High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome. Nature *494*, 385–389.

Hashem, Y. et al. (2013b). Structure of the Mammalian Ribosomal 43S Preinitiation Complex Bound to the Scanning Factor DHX29. Cell *153*, 1108–1119.

Henras, A.K. et al. (2008). The post-transcriptional steps of eukaryotic ribosome biogenesis. Cell. Mol. Life Sci. *65*, 2334–2359.

Hiller, D.A. et al. (2011). A two-step chemical mechanism for ribosome-catalysed peptide bond formation. Nature *476*, 236–239.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb. Perspect. Biol. *4*, a011544.

Houge, G. et al. (1993). Selective cleavage of 28S rRNA variable regions V3 and V13 in myeloid leukemia cell apoptosis. FEBS Lett. *315*, 16–20.

Houge, G. et al. (1995). Fine mapping of 28S rRNA sites specifically cleaved in cells undergoing apoptosis. Mol. Cell. Biol. *15*, 2051–2062.

Hsiao, C. et al. (2009). Peeling the onion: ribosomes are ancient molecular fossils. Mol. Biol. Evol. *26*, 2415–2425.

Hury, J. et al. (2006). Ribosome origins: the relative age of 23S rRNA Domains. Orig. Life Evol. Biosph. *36*, 421–429.

Jackson, R.J. et al. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jackson, R.J. et al. (2012). Termination and post-termination events in eukaryotic translation. Adv. Protein Chem. Struct. Biol. *86*, 45–93.

Jeeninga, R.E. et al. (1997). Variable regions V13 and V3 of Saccharomyces cerevisiae contain structural features essential for normal biogenesis and stability of 5.8S and 25S rRNA. RNA *3*, 476–488.

Jenner, L. et al. (2010). Structural rearrangements of the ribosome at the tRNA proofreading step. Nat. Struct. Mol. Biol. *17*, 1072–1078.

Jenner, L. et al. (2012). Crystal structure of the 80S yeast ribosome. Curr. Opin. Struct. Biol. *22*, 759–767.

Jorgensen, R. et al. (2006). The life and death of translation elongation factor 2. Biochem. Soc. Trans. *34*, 1–6.

Kaminishi, T. et al. (2007). A snapshot of the 30S ribosomal subunit capturing mRNA via the Shine-Dalgarno interaction. Structure *15*, 289–297.

Kavran, J.M., and Steitz, T.A. (2007). Structure of the base of the L7/L12 stalk of the Haloarcula marismortui large ribosomal subunit: analysis of L11 movements. J. Mol. Biol. *371*, 1047–1059.

Kingery, D.A. et al. (2008). An uncharged amine in the transition state of the ribosomal peptidyl transfer reaction. Chem. Biol. *15*, 493–500.

Klaholz, B.P. (2011). Molecular recognition and catalysis in translation termination complexes. Trends Biochem. Sci. *36*, 282–292.

Klein, D.J. et al. (2001). The kink-turn: a new RNA secondary structure motif. EMBO J. *20*, 4214–4221.

Klein, D.J. et al. (2004). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. J. Mol. Biol. *340*, 141–177.

Klinge, S. et al. (2011). Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. Science *334*, 941–948.

Klinge, S. et al. (2012). Atomic structures of the eukaryotic ribosome. Trends Biochem. Sci. *3*7, 189–198.

Kondrashov, N. et al. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell *145*, 383–397.

Konevega, A.L. et al. (2007). Spontaneous reverse movement of mRNA-bound tRNA through the ribosome. Nat. Struct. Mol. Biol. *14*, 318–324.

Korostelev, A. et al. (2008). Crystal structure of a translation termination complex formed with release factor RF2. Proc. Natl. Acad. Sci. U.S.A. *105*, 19684–19689.

Kressler, D. et al. (2010). Driving ribosome assembly. Biochim. Biophys. Acta *1803*, 673–683.

Kuhlenkoetter, S. et al. (2011). Different substrate-dependent transition states in the active site of the ribosome. Nature *476*, 351–354.

Lafontaine, D.L., and Tollervey, D. (2001). The function and synthesis of ribosomes. Nat. Rev. Mol. Cell. Biol. *2*, 514–520.

Larsson, S.L., and Nygård, O. (2001). Proposed secondary structure of eukaryote specific expansion segment 15 in 28S rRNA from mice, rats, and rabbits. Biochemistry *40*, 3222–3231.

Laurberg, M. et al. (2008). Structural basis for translation termination on the 70S ribosome. Nature *454*, 852–857.

Lecompte, O. et al. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res. *30*, 5382–5390.

Lee, A.S.-Y. et al. (2013). A ribosome-specialized translation initiation pathway is required for cap-dependent translation of vesicular stomatitis virus mRNAs. Proc. Natl. Acad. Sci. U.S.A. *110*, 324–329.

Leidig, C. et al. (2013). Structural characterization of a eukaryotic chaperone-the ribosome-associated complex. Nat. Struct. Mol. Biol. *20*, 23–28.

Leung, E.K.Y. et al. (2011). The mechanism of peptidyl transfer catalysis by the ribosome. Annu. Rev. Biochem. *80*, 527–555.

Li, X. et al. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods *10*, 584–590.

Lomakin, I.B., and Steitz, T.A. (2013). The initiation of mammalian protein synthesis and mRNA scanning mechanism. Nature *500*, 307-311.

Lutsch, G. et al. (1990). Immunoelectron microscopic studies on the location of ribosomal proteins on the surface of the 40S ribosomal subunit from rat liver. Eur. J. Cell Biol. *51*, 140–150.

Marion, M.J., and Marion, C. (1987). Localization of ribosomal proteins on the surface of mammalian 60S ribosomal subunits by means of immobilized enzymes. Correlation with chemical cross-linking data. Biochem. Biophys. Res. Commun. *149*, 1077–1083.

Márquez, V. et al. (2011). Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. J. Mol. Biol. *405*, 1215–1232.

Mateyak, M.K., and Kinzy, T.G. (2013). ADP-ribosylation of Translation Elongation Factor 2 by Diphtheria Toxin in Yeast Inhibits Translation and Cell Separation. J. Biol. Chem. *288*, 24647–24655.

Mears, J.A. et al. (2002). Modeling a minimal ribosome based on comparative sequence analysis. J. Mol. Biol. *321*, 215–234.

Melnikov, S. et al. (2012). One core, two shells: bacterial and eukaryotic ribosomes. Nat. Struct. Mol. Biol. *19*, 560–567.

Ménétret, J.F. et al. (2000). The structure of ribosome-channel complexes engaged in protein translocation. Mol. Cell *6*, 1219–1232.

Meyuhas, O., and Dreazen, A. (2009). Ribosomal protein S6 kinase from TOP mRNAs to cell size. Prog. Mol. Biol. Transl. Sci. *90*, 109–153.

Morgan, D.G. et al. (2000). A comparison of the yeast and rabbit 80 S ribosome reveals the topology of the nascent chain exit tunnel, inter-subunit bridges and mammalian rRNA expansion segments. J. Mol. Biol. *301*, 301–321.

Nilsson, J. et al. (2004). Regulation of eukaryotic translation by the RACK1 protein: a platform for signalling molecules on the ribosome. EMBO Rep. *5*, 1137–1141.

Nissen, P. et al. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science *289*, 920–930.

Nissen, P. et al. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. Proc. Natl. Acad. Sci. U.S.A. *98*, 4899–4903.

Noller, H.F. (2005). RNA structure: reading the ribosome. Science *309*, 1508–1514.

Noller, H.F. (2012). Evolution of protein synthesis from an RNA world. Cold Spring Harb. Perspect. Biol. *4*, a003681.

Nomura, M. (1973). Assembly of bacterial ribosomes. Science *179*, 864–873.

Nygård, O. et al. (2006). Analysis of the secondary structure of expansion segment 39 in ribosomes from fungi, plants and mammals. J. Mol. Biol. *357*, 904–916.

Ogle, J.M. et al. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science *292*, 897–902.

Ogle, J.M. et al. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell *111*, 721–732.

Ogle, J.M., and Ramakrishnan, V. (2005). Structural insights into translational fidelity. Annu. Rev. Biochem. *74*, 129–177.

Palade, G.E. (1955). A small particulate component of the cytoplasm. J. Biophys. Biochem. Cytol. *1*, 59–68.

Panse, V.G., and Johnson, A.W. (2010). Maturation of eukaryotic ribosomes: acquisition of functionality. Trends Biochem. Sci. *35*, 260–266.

Petry, S. et al. (2005). Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. Cell *123*, 1255–1266.

Petry, S. et al. (2008). The termination of translation. Curr. Opin. Struct. Biol. *18*, 70–77.

Pfeffer, S. et al. (2012). Structure and 3D arrangement of endoplasmic reticulum membrane-associated ribosomes. Structure *20*, 1508–1518.

Pisarev, A.V. et al. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. Mol. Cell *37*, 196–210.

Pomeranz Krummel, D.A. et al. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. Nature *458*, 475–480.

Prokopowich, C.D. et al. (2003). The correlation between rDNA copy number and genome size in eukaryotes. Genome *46*, 48–50.

Pulk, A., and Cate, J.H.D. (2013). Control of ribosomal subunit rotation by elongation factor G. Science *340*, 1235970.

Rabl, J. et al. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. Science *331*, 730–736.

Ratje, A.H. et al. (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature *468*, 713–716.

Richard, G.-F. et al. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol. Mol. Biol. Rev. *72*, 686–727.

Richter, J.D., and Lasko, P. (2011). Translational control in oocyte development. Cold Spring Harb. Perspect. Biol. *3*, a002758.

Rodnina, M.V., and Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. Annu. Rev. Biochem. *70*, 415–435.

Rodnina, M.V., and Wintermeyer, W. (2009). Recent mechanistic insights into eukaryotic ribosomes. Curr. Opin. Cell Biol. *21*, 435–443.

Ruvinsky, I., and Meyuhas, O. (2006). Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. Trends Biochem. Sci. *31*, 342–348.

Schluenzen, F. et al. (2000). Structure of functionally activated small ribosomal subunit at 3.3 ångstroms resolution. Cell *102*, 615–623.

Schmeing, T.M. et al. (2005). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. Nature *438*, 520–524.

Schmeing, T.M. et al. (2009). The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. Science *326*, 688–694.

Schmeing, T.M. et al. (2011). How mutations in tRNA distant from the anticodon affect the fidelity of decoding. Nat. Struct. Mol. Biol. *18*, 432–436.

Schmeing, T.M., and Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. Nature *461*, 1234–1242.

Schuette, J.-C. et al. (2009). GTPase activation of elongation factor EF-Tu by the ribosome during decoding. EMBO J. *28*, 755–765.

Schuwirth, B.S. et al. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. Science *310*, 827–834.

Selmer, M. et al. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. Science *313*, 1935–1942.

Sengupta, J. et al. (2001). Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. Proc. Natl. Acad. Sci. U.S.A. *98*, 11991–11996.

Sengupta, J. et al. (2004). Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM. Nat. Struct. Mol. Biol. *11*, 957–962.

Sengupta, J. et al. (2010). Characterization of the nuclear export adaptor protein Nmd3 in association with the 60S ribosomal subunit. J. Cell Biol. *189*, 1079–1086.

Shajani, Z. et al. (2011). Assembly of bacterial ribosomes. Annu. Rev. Biochem. *80*, 501–526.

Sharma, G. et al. (2012). Affinity grid-based cryo-EM of PKC binding to RACK1 on the ribosome. J. Struct. Biol. 1–5.

Shasmal, M., and Sengupta, J. (2012). Structural Diversity in Bacterial Ribosomes: Mycobacterial 70S Ribosome Structure Reveals Novel Features. PLoS ONE *7*, e31742.

Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. *71*, 1342–1346.

Shoji, S. et al. (2006). Reverse translocation of tRNA in the ribosome. Mol. Cell *24*, 931–942.

Sievers, A. et al. (2004). The ribosome as an entropy trap. Proc. Natl. Acad. Sci. U.S.A. *101*, 7897–7901.

Simonetti, A. et al. (2009). A structural view of translation initiation in bacteria. Cell. Mol. Life Sci. *66*, 423–436.

Spahn, C.M. et al. (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions. Cell *107*, 373–386.

Spahn, C.M.T. et al. (2004a). Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. EMBO J. *23*, 1008–1019.

Spahn, C.M.T. et al. (2004b). Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. Cell *118*, 465–475.

Stark, H. et al. (1997). Visualization of elongation factor Tu on the *Escherichia coli* ribosome. Nature *389*, 403–406.

Steitz, T.A. (2008). A structural understanding of the dynamic ribosome machine. Nat. Rev. Mol. Cell. Biol. *9*, 242–253.

Strunk, B.S. et al. (2011). Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. Science *333*, 1449–1453.

Sweeney, R. et al. (1994). An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. Mol. Cell. Biol. *14*, 4203–4215.

Tafforeau, L. et al. (2013). The Complexity of Human Ribosome Biogenesis Revealed by Systematic Nucleolar Screening of Pre-rRNA Processing Factors. Mol. Cell *51*, 539–551.

Takahashi, Y. et al. (2002). Identification of the ribosomal proteins present in the vicinity of globin mRNA in the 40S initiation complex. J. Biochem. *132*, 705–711.

Tautz, D. et al. (1988). Complete sequences of the rRNA genes of *Drosophila melanogaster*. Mol. Biol. Evol. *5*, 366–376.

Taylor, D.J. et al. (2007). Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. EMBO J. *26*, 2421–2431.

Taylor, D.J. et al. (2009). Comprehensive Molecular Structure of the Eukaryotic Ribosome. Structure *17*, 1591–1604.

Tourigny, D.S. et al. (2013). Elongation factor G bound to the ribosome in an intermediate state of translocation. Science *340*, 1235490.

Triana-Alonso, F.J. et al. (1995). The elongation factor 3 unique in higher fungi and essential for protein biosynthesis is an E site factor. J. Biol. Chem. *270*, 20473–20478.

Ude, S. et al. (2013). Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science *339*, 82–85.

Verschoor, A. et al. (1996). Native 3D structure of eukaryotic 80s ribosome: morphological homology with *E. coli* 70S ribosome. J. Cell Biol. *133*, 495–505.

Voigts-Hoffmann, F. et al. (2012). Structural insights into eukaryotic ribosomes and the initiation of translation. Curr. Opin. Struct. Biol. *22*, 768–777.

Voorhees, R.M. et al. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat. Struct. Mol. Biol. *16*, 528–533.

Voorhees, R.M. et al. (2010). The mechanism for activation of GTP hydrolysis on the ribosome. Science *330*, 835–838.

Voorhees, R.M., and Ramakrishnan, V. (2013). Structural basis of the translational elongation cycle. Annu. Rev. Biochem. *82*, 203–236.

Wang, D.O. et al. (2010). Spatially restricting gene expression by local translation at synapses. Trends Neurosci. *33*, 173–182.

Wang, X. et al. (2002). Modular recognition of RNA by a human pumilio-homology domain. Cell *110*, 501–512.

Ware, V.C. et al. (1985). rRNA processing: removal of only nineteen bases at the gap between 28S alpha and 28S beta rRNAs in *Sciara coprophila*. Nucleic Acids Res. *13*, 3581–3597.

Weisser, M. et al. (2013). The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. Nat. Struct. Mol. Biol. *20*, 1015–1017.

Weixlbaumer, A. et al. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. Science *322*, 953–956.

Wild, T. et al. (2010). A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60S subunit export. PLoS Biol. *8*, e1000522.

Wilson, D.N. (2009). The A-Z of bacterial translation inhibitors. Crit. Rev. Biochem. Mol. Biol. *44*, 393–433.

Wilson, D.N., and Cate, J.H.D. (2012). The structure and function of the eukaryotic ribosome. Cold Spring Harb. Perspect. Biol. *4*, a011536.

Wilson, D.N., and Nierhaus, K.H. (2005). Ribosomal proteins in the spotlight. Crit. Rev. Biochem. Mol. Biol. *40*, 243–267.

Wimberly, B.T. et al. (2000). Structure of the 30S ribosomal subunit. Nature *407*, 327–339.

Xue, S., and Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat. Rev. Mol. Cell. Biol. *13*, 355–369.

Yokoyama, T., and Suzuki, T. (2008). Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments. Nucleic Acids Res. *36*, 3539–3551.

Yu, Y. et al. (2011). Common conformational changes induced in type 2 picornavirus IRESs by cognate trans-acting factors. Nucleic Acids Res. *39*, 4851–4865.

Yusupov, M.M. et al. (2001). Crystal structure of the ribosome at 5.5 Å resolution. Science *292*, 883–896.

Yutin, N. et al. (2012). Phylogenomics of prokaryotic ribosomal proteins. PLoS ONE *7*, e36972.

Zaher, H.S., and Green, R. (2009). Fidelity at the molecular level: lessons from protein synthesis. Cell *136*, 746–762.

Zavialov, A.V. et al. (2001). A posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. Cell *107*, 115–124.

Zavialov, A.V. et al. (2005). Splitting of the posttermination ribosome into subunits by the concerted action of RRF and EF-G. Mol. Cell *18*, 675–686.

Zhang, Y., and Lu, H. (2009). Signaling to p53: ribosomal proteins find their way. Cancer Cell *16*, 369–377.

Zhou, J. et al. (2013). Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation. Science *340*, 1236086.

7 Publications

2013

1) **Anger AM***, Armache JP*, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R (2013). Structures of the human and *Drosophila* 80S ribosome. ***Nature***, 497(7447), 80-85.

2) Armache JP*, **Anger AM*,** Márquez V, Franckenberg S, Fröhlich T, Villa E, Berninghausen O, Thomm M, Arnold GJ, Beckmann R, Wilson DN (2013). Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. ***Nucl. Acids Res.***, 41(2), 1284-1293.

2012

3) Becker T, Franckenberg S, Wickles S, Shoemaker CJ, **Anger AM**, Armache JP, Sieber H, Ungewickell C, Berninghausen O, Daberkow I, Karcher A, Thomm M, Hopfner KP, Green R, Beckmann R (2012). Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. ***Nature***, *482*(7386), 501–506.

2011

4) Becker T, Armache JP, Jarasch A, **Anger, AM**, Villa E, Sieber H, Abdel Motaal B, Mielke T, Berninghausen O, Beckmann R (2011). Structure of the no-go mRNA decay complex Dom34–Hbs1 bound to a stalled 80S ribosome. ***Nat. Struct. Mol. Biol.***, *18*(6), 715–720.

2010

5) Armache JP*, Jarasch A*, **Anger AM***, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Söding J, Westhof E, Wilson DN, Beckmann R (2010). Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. ***Proc. Natl. Acad. Sci. USA***, *107*(46), 19754–19759.

6) Armache JP*, Jarasch A*, **Anger AM***, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Söding J, Westhof E, Wilson DN, Beckmann R (2010). Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. ***Proc. Natl. Acad. Sci. USA***, *107*(46), 19748–19753.

*denotes equal contribution.

1284–1293 *Nucleic Acids Research, 2013, Vol. 41, No. 2* *Published online 6 December 2012*
doi:10.1093/nar/gks1259

Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution

Jean-Paul Armache[1,2], Andreas M. Anger[1,2], Viter Márquez[1,2], Sibylle Franckenberg[1,2], Thomas Fröhlich[1,3], Elizabeth Villa[4], Otto Berninghausen[1,2], Michael Thomm[5], Georg J. Arnold[1,3], Roland Beckmann[1,2,*] and Daniel N. Wilson[1,2,*]

[1]Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany, [2]Center for integrated Protein Science Munich (CiPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany, [3]Laboratory for Functional Genome Analysis LAFUGA, Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor-Lynen-Str. 25, 81377 Munich, Germany, [4]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany and [5]Lehrstuhl für Mikrobiologie, University of Regensburg, Universitätstraße 32, 93053 Regensburg, Germany

Received September 27, 2012; Revised and Accepted November 1, 2012

ABSTRACT

In all living cells, protein synthesis occurs on ribonucleoprotein particles called ribosomes. Molecular models have been reported for complete bacterial 70S and eukaryotic 80S ribosomes; however, only molecular models of large 50S subunits have been reported for archaea. Here, we present a complete molecular model for the *Pyrococcus furiosus* 70S ribosome based on a 6.6 Å cryo-electron microscopy map. Moreover, we have determined cryo-electron microscopy reconstructions of the Euryarchaeota *Methanococcus igneus* and *Thermococcus kodakaraensis* 70S ribosomes and Crenarchaeota *Staphylothermus marinus* 50S subunit. Examination of these structures reveals a surprising promiscuous behavior of archaeal ribosomal proteins: We observe intersubunit promiscuity of S24e and L8e (L7ae), the latter binding to the head of the small subunit, analogous to S12e in eukaryotes. Moreover, L8e and L14e exhibit intrasubunit promiscuity, being present in two copies per archaeal 50S subunit, with the additional binding site of L14e analogous to the related eukaryotic r-protein L27e. Collectively, these findings suggest insights into the evolution of eukaryotic ribosomal proteins through increased copy number and binding site promiscuity.

INTRODUCTION

In all three domains of life, protein synthesis in the cell is performed by large macromolecular machines called ribosomes (1–3). In bacteria, such as *Escherichia coli*, the 70S ribosome is formed from a small 30S subunit, comprising one 16S ribosomal RNA (rRNA) and 21 ribosomal proteins (r-proteins), and a large 50S subunit composed of a 5S and 23S rRNA and 33 r-proteins (3). X-ray structures of bacterial ribosomal particles have revealed a complex network of interactions between the rRNAs and r-proteins (4–6). In comparison, eukaryotic 80S ribosomes are larger and more complex than bacterial ribosomes. For example, the yeast small 40S subunit contains one 18S rRNA and 33 r-proteins, whereas the large 60S subunit comprises 5S, 5.8S and 25S rRNAs together with 46 large subunit r-proteins (3). Crystal structures of the *Saccharomyces cerevisiae* 80S ribosome (7) and *Tetrahymena thermophila* 40S and 60S subunits (8,9) have determined the architecture of the additional rRNA expansion segments (ES) and variable regions (VR) as well as the localization of the eukaryotic-specific r-proteins.

Genomic studies indicate that archaeal ribosomes have an intermediate complexity compared with bacterial and eukaryotic ones (10–12). Although a 6.6 Å cryo-electron microscopy (EM) map of the archaeal 70S ribosome exists, no molecular model was reported (13). So far, the large 50S subunit from the Euryarchaeota *Haloarcula marismortui* has been crystallized disclosing

*To whom correspondence should be addressed. Tel: +49 89 2180 76903; Fax: +49 89 2180 76999; Email: wilson@lmb.uni-muenchen.de
Correspondence may also be addressed to Roland Beckmann. Tel: +49 89 2180 76900; Fax: +49 89 2180 76945;
Email: beckmann@lmb.uni-muenchen.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

structures for the 5S and 23S rRNA as well as 27 r-proteins, 12 of which are archaeal/eukaryotic-specific (14). Recently, a cryo-EM structure of the Euryarchaeota *Methanobacterium thermoautotrophicus* 50S subunit was determined, leading to the identification of five additional archaeal/eukaryotic specific r-proteins and some rRNA ES that are not present in the *H. marismortui* 50S subunit structure (15). Here, we present the complete molecular model for the *Pyrococcus furiosus* 70S ribosome, using a 6.6 Å cryo-EM structure (13). Based on 2D-PAGE and mass spectrometry (MS) analysis of *Thermococcus kodakaraensis* ribosome, coupled with additional cryo-EM reconstructions of Euryarchaeota *Methanococcus igneus* and *T. kodakaraensis* 70S ribosomes, and Crenarchaeota *Staphylothermus marinus* 50S subunit, we reveal a surprising promiscuity of r-proteins within archaeal ribosomes that has implications for the evolution of r-proteins in archaea and eukaryotes.

MATERIALS AND METHODS

Growth of *M. igneus*, *S. marinus* and *T. kodakaraensis*

T. kodakaraensis KOD1^T (JCM 12380^T, ATCC BAA-918^T), *M. igneus* Kol5^T (DSM 5666) and *S. marinus* F1^T (DSM 3639, ATCC 49053^T) were obtained from the culture collection of the Institute of Microbiology and Archaea Centre, University of Regensburg. *T. kodakaraensis* and *S. marinus* were grown under anoxic conditions in Marine-*Thermococcus*-medium (16) at 85°C and pH 7.0. Substrates 0.1% yeast extract and 0.1% peptone were added. For the cultivation of *S. marinus*, the medium was further supplemented with 0.7% elemental sulfur. The gas phases consisted of N_2/CO_2 (80/20 v/v, 250 kPa). *M. igneus* was grown in *Methanotorris*-medium at 85°C as previously described (17). The *Methanotorris*-medium consisted (per liter) of K_2HPO_4 0.0556 g, KH_2PO_4 0.0558 g, KCl 0.269 g, NaCl 25.14 g, $NaHCO_3$ 0.84 g, $CaCl_2 \times 2H_2O$ 0.368 g, $MgCl_2 \times 6H_2O$ 7.724 g, NH_4Cl 1.180 g, $Fe(NH_4)_2(SO_4)_2 \times 6H_2O$ 12 mg, Na_3-Nitrilotriacetate (Titriplex I) 88 mg, Na_2SeO_4 8 mg, $Na_2WO_4 \times 2H_2O$ 3.2 mg, $Na_2MoO_4 \times 2H_2O$ 2.4 mg and 10 ml of trace element solution (18). The medium was covered with a gas phase of H_2/CO_2 (80/20 v/v, 250 kPa) and chemically reduced with Na_2S x $9H_2O$ (0.4 g/l). Mass cultivations for all strains were carried out in 300l enamel-protected fermenters (HTE, Bioengineering, Wald, Switzerland) with 250l culture medium and continuous supply of 0.1% sodium sulfide for pH stabilization. The cells were harvested by centrifugation (Padberg, Lahr, Germany), shock-frozen in liquid nitrogen and stored at −80°C until further use.

Preparation of ribosomal particles

Archaeal ribosomes were isolated and purified in a similar fashion as previously reported (19,20); however, with some modifications as described in (21): Cell pellets were dissolved in Tico buffer (20 mM Hepes pH 7.5, 10 mM $Mg(OAc)_2$, 30 mM NH_4OAc, 4 mM ß-Mercaptoethanol) at 4°C and subsequently disrupted by using a Microfluidizer (Microfluidics M-110L Pneumatic) at 18 000 psi. The crude homogenate was centrifuged twice at 30 000*g* at 4°C for 30 min to obtain the S30 fraction. A crude ribosomal fraction was obtained by centrifugation at 100 000*g* for 5 h at 4°C and dissolving the pellet in an equal volume of high salt wash (HSW) buffer (20 mM Hepes, 10 mM $Mg(OAc)_2$, 500 mM NH_4OAc, 4 mM ß-Mercaptoethanol, pH 7.5). Large debris were removed by centrifuging the crude ribosomes for 5 min at 18 000*g* at 4°C. The clear supernatant was diluted 10-fold in HSW buffer and layered on top of 1.3 volumes of 25% (w/v) sucrose cushion prepared in HSW buffer and centrifuged at 100 000*g* for 7 h at 4°C. The pellet was resuspended in a minimal volume of Tico buffer and subsequently purified using a sucrose-density gradient centrifugation (10–40% sucrose in Tico buffer) at 46 000*g* for 17 h at 4°C. Fractions corresponding to the 50S and 30S were separately pooled and pelleted at 140 000*g* for 12 h at 4°C and resuspended in a minimal volume of Tico buffer.

Extraction of the total ribosomal proteins, 2D-PAGE and MS

The total proteins from HSW ribosomes and purified ribosomal subunits were extracted by acetic acid according to Nierhaus *et al.* (22). Lyophilized proteins were further processed for liquid chromatography tandem MS (LC-MS/MS) analysis and for 2D-PAGE. Around 2 µg of total proteins was necessary for the LC-MS/MS, whereas 5–10 µg of total proteins were required for 2D-PAGE. The 2D electrophoresis was performed as described by Kaltschmidt and Wittman (23). LC-MS/MS analysis of 2D-gel spots on the LTQ ion-trap and of protein samples on the Orbitrap XL instrument was performed as previously described (21).

Electron microscopy

Cryo-EM and single particle reconstruction

As described previously (24), *M. igneus* and *T. kodakaraensis* 70S ribosomes and *S. marinus* 50S subunits were applied to carbon-coated holey grids. Images were collected on a Tecnai G2 Spirit TEM at 120 kV at a nominal magnification of 90 000 using a 2 K Eagle (2048 × 2048) CCD camera (FEI) resulting in a pixel size of 3.31 Å/pixel. The data were analyzed by determination of the contrast transfer function using CTFFIND (25), and further processed with the SPIDER software package (26), using the *H. marismortui* 50S subunit (PDB3CC2) (14) filtered to between 20–25 Å, as an initial reference. Further steps involved subsequent refinement and iterative sorting for heterogeneities. For the final reconstructions, 8932 particles were used for *M. igneus*, 10 431 for *T. kodakaraensis* and 11 142 for *S. marinus*; this resulted in 18 Å, 19 Å and 24 Å electron density maps at 0.5 FSC, respectively.

Modeling and figure generation

P. furiosus rRNA modeling

P. furiosus 16S and 23S/5S rRNA sequences were taken from GeneBank Accession number (Acc.) U20163 and Acc. AE009950, respectively. Structure-based sequence

alignments of the conserved rRNA core were constructed using Sequence to Structure (S2S) (27) based on X-ray structures of the small ribosomal subunit of *Thermus thermophilus* for the 16S rRNA [Protein Data Bank (PDB) code 1fjf] (28) and the large ribosomal subunit of *H. marismortui* for the 23S and 5S rRNAs (PDB 1ffk) (14). For regions like the stalk base (H42–H44), H69, the L1 stalk (H76–H79) and the base of H98 (ES39L), the corresponding structures of the large ribosomal subunit of *T. thermophilus* and *Escherichia coli* (PDBs $2 \times 9u$ and 3r8s) (29,30) were used as templates. Two eukaryote-like rRNA parts (h33 and H54–H59) were generated in separate S2S alignments based on the X-ray structure of the 80S ribosome from *S. cerevisiae* (PDBs 3o58 and 3o2z) (7). All remaining parts of the structure were built *de novo* using Assemble (31) essentially as described (32). The resulting complete models of the *P. furiosus* rRNAs were manually inspected and adjusted according to features of the electron density using Coot (33).

P. furiosus r-protein modeling

We used a 6.6 Å cryo-EM map of *P. furious* 70S ribosome (13) to localize and build models for 62 archaeal proteins (66 when including L8e(2), L8e(S), S24e(L) and L14e(2)). This consists of 33 r-proteins common to all three domains, 30 archaea/eukaryote-specific r-proteins (if LX is considered as L20e). A total of 27 proteins from the 50S subunit were modeled using archaeal X-ray *H. marismortui* protein templates (PDB 3cc2). In all, 25 r-proteins from the 30S subunit and five proteins from the 50S subunit (L14e, L34e, L35ae, L40e and L41e) were modeled using templates from the eukaryotic *S. cerevisiae* X-ray structure (PDBs 3u5g and 3u5c for the small subunit, PDBs 3u5e and 3u5i for the large subunit). Protein LX was modeled using *Methanobacterium thermoautothropicus* PDB 2jxt; the stalk protein P0 was based on *Pyrococcus horikoshii* and *Methanocaldococcus janaschii*, PDBs 3a1y and 3jsy, respectively; the L1 protein was based on *T. thermophilus* (Bacteria) PDB 2hw8 template. The multi-sequence alignment was performed using ClustalW (34), whereas for sequence analysis, Jalview was used. Protein models were created using Modeller (35) and further fit and analyzed using Chimera (36) (rigid body fitting) and Coot (33) (manual adjustments), as well as MDFF (37) in VMD (38).

Refinement and fitting of the rRNA and r-proteins into the EM densities

MDFF was used to refine the proteins and RNA into the density while fixing protein–RNA and protein–protein clashes, followed by an MDFF refinement of the entire 70S model.

Figure generation

Figures showing electron densities and atomic models were generated using Chimera (36).

RESULTS

A cryo-EM map and model for the archaeal 70S ribosome

In contrast to purifications of archaeal ribosomal particles from Crenarchaeota *Sulfolobus acidocaldarius*, *Pyrobaculum aerophilum* (19,21), *S. marinus* (Figure 1A, upper panel) and Euryarchaeota *M. thermoautotrophicus* (15) where only 30S and 50S subunits were obtained, we were able to isolate intact archaeal 70S ribosomes from Euryarchaeota *P. furiosus* and *T. kodakaraensis* translation extracts using sucrose gradient centrifugation (Figure 1A, lower panel). A cryo-EM map of the *P. furiosus* 70S ribosome at 6.6 Å (0.5 FSC) (Figure 1B) (13) was then used to generate a molecular model for the

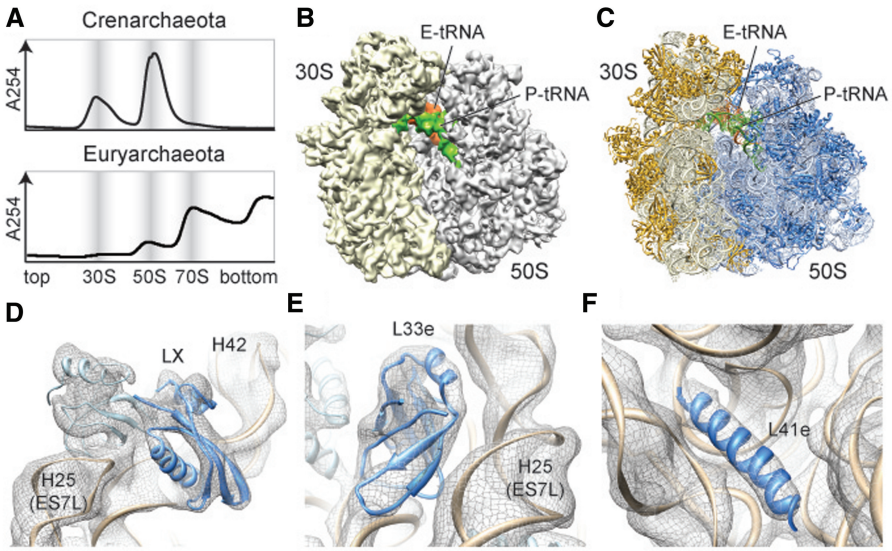


Figure 1. Cryo-EM structure and molecular model of an archaeal 70S ribosome. (**A**) Sucrose density gradient centrifugation profile of ribosomal particles from the Crenarcheaota *S. marinus* (Upper panel) and Euryarchaeota *T. kodakaraensis* (lower panel), with 30S, 50S and 70S peaks highlighted. (**B**) Cryo-EM reconstruction (30S, yellow; 50S, gray) and (**C**) molecular model (16S and 23 rRNA, light yellow and blue; SSU and LSU r-proteins, gold and blue) of the *P. furiosus* 70S ribosome. P- and E-tRNA are colored green and orange, respectively. (**D–F**) Fit of molecular models for rRNA (tan) and r-proteins (blue), (D) LX, (E) L33e and (F) L41e, into the cryo-EM density of the *P. furiosus* 70S ribosome (gray mesh).

rRNA and r-protein components (Figure 1C). The *P. furiosus* rRNAs were built in S2S (27) and Assemble (31), using initial models based on templates derived from the X-ray structures of the bacterial 30S (28) and 70S (29,30), the archaeal 50S (14) and the eukaryotic 80S (7) (see Materials and methods for details). Thus, in addition to the conserved rRNA core, five variable regions [VR5S (h16-h17), VR8S (h33), VR1L (H1), VR22L (H58) and H16-18] and 12 ES (ES7S, ES9S, ES4L, ES5L, ES7L, ES9L, ES15L, ES20L, ES24L, ES26L, ES39L and ES41L) were modeled (Supplementary Figure S1). The majority of the *P. furiosus* VRs and ESs had conformations remarkably similar to the equivalent regions in the eukaryotic ribosome (7–9,32) (Supplementary Figure S2), as noted previously for the *M. thermoautotrophicus* 50S subunit (15). However, VR5S (h16), VR1L (H1) and ES39L adopt novel conformations in the *P. furiosus* 70S (Supplementary Figure S3) that have not been observed in previous ribosome structures. Unlike the *M. thermoautotrophicus* 23S rRNA, which is a composite of *H. marismortui* and *T. thermophila* rRNA fragments (15), we present complete models for the *P. furiosus* 16S and 23S rRNAs (Supplementary Figure S4) with continuous *P. furiosus* sequence and numbering (Supplementary Figures S1, S5–S7).

Based on genomic analysis, the *P. furiosus* 70S ribosome is predicted to contain 64 r-proteins, 25 in the 30S and 39 in the 50S subunit (10–12) (Supplementary Tables S1–S4). In all, 35 (30S, 15; 50S, 20) of the *P. furiosus* r-proteins have counterparts in bacteria, and the location of an additional 12 large subunit r-proteins is known from the X-ray structure of the *H. marismortui* 50S subunit (14). Locations for the remaining 17 (30S, 10; 50S, 7) were determined by homology with the respective eukaryotic r-proteins present in the X-ray structure of the *S. cerevisiae* 80S ribosome (7). The models for *P. furiosus* r-proteins L14e [we use the revised and simplified nomenclature based on family names for eukaryotic r-proteins (7), see Supplementary Tables S1–S4], L30e, L34e, L40e and LX (Figure 1D) were in agreement with those reported recently for the euryarchaeotal *M. thermoautotrophicus* 50S subunit (15). In addition to 10 models of small subunit *P. furiosus* r-proteins, we also present models for r-proteins L33e (Figure 1E) and L41e (Figure 1F), which are absent in the genomes of *H. marismortui* and *M. thermoautotrophicus* (Supplementary Tables S1 and S2) (10–12). The high quality of the *P. furiosus* 70S ribosome cryo-EM map enabled an accurate fit of the molecular models of the rRNA and r-proteins by using distinct features of the electron density seen for the major and minor grooves of the RNA helices and rod-like densities for r-proteins (Supplementary Figure S4). Surprisingly, after fitting all rRNA and r-protein, four regions of additional density remained unaccounted for; one located on

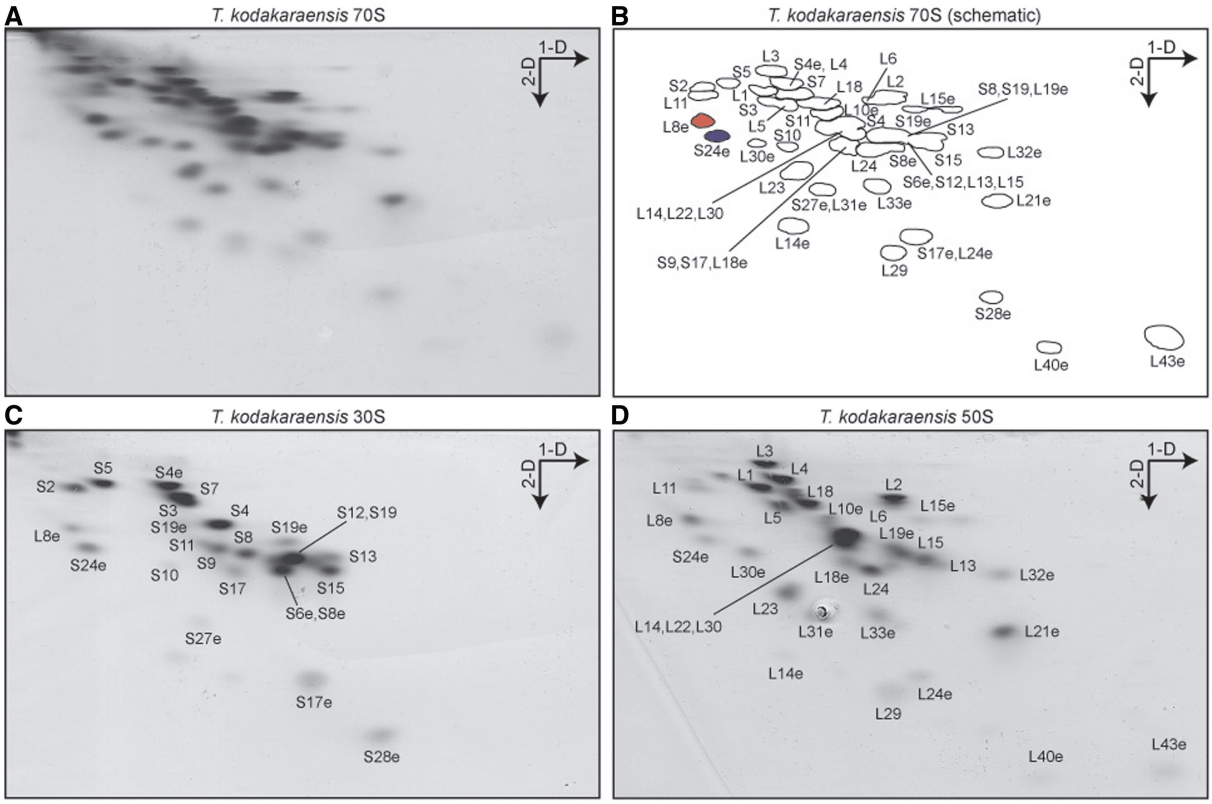


Figure 2. 2D-PAGE and MS analysis of Euryarchaeota ribosomal proteins. (**A–D**) Coomassie blue-stained 2D gel of (A) high-salt-washed *T. kodakaraensis* 70S ribosomes (MS assignments of protein spots labeled in (B) scheme), and sucrose gradient purified (C) 30S and (D) 50S subunits. The direction of the first (1D, based on charge) and second (2D, based on mass) dimensions of electrophoresis are indicated with arrows, and spots for L8e (L7ae) and S24e are colored.

the 30S subunit and three on the 50S subunit. As all the rRNA was satisfactorily modeled and the additional density had features reminiscent of protein, we hypothesized that the additional density was owing to the presence of yet unidentified r-proteins.

2D-PAGE and MS of Euryarchaeota ribosomes

Our previous MS analysis of the Crenarchaeota *S. acidocaldarius* and *P. aerophilum* ribosomes led to the identification of a number of hypothetical proteins with basic isoelectric point that were ribosome associated (21). Thus, to search for additional r-proteins in Euryarchaeota, we performed 2D-PAGE (23) on high-salt washed 70S ribosomes from *T. kodakaraensis* (previously called *Pyrococcus kodakaraensis*) (Figure 2A and B), which belongs to the same *Thermococcaceae* family as *P. furiosus*. MS identification of the protein spots, coupled with LC-MS/MS analysis of the total protein samples, led to the identification of 25/25 (100%) and

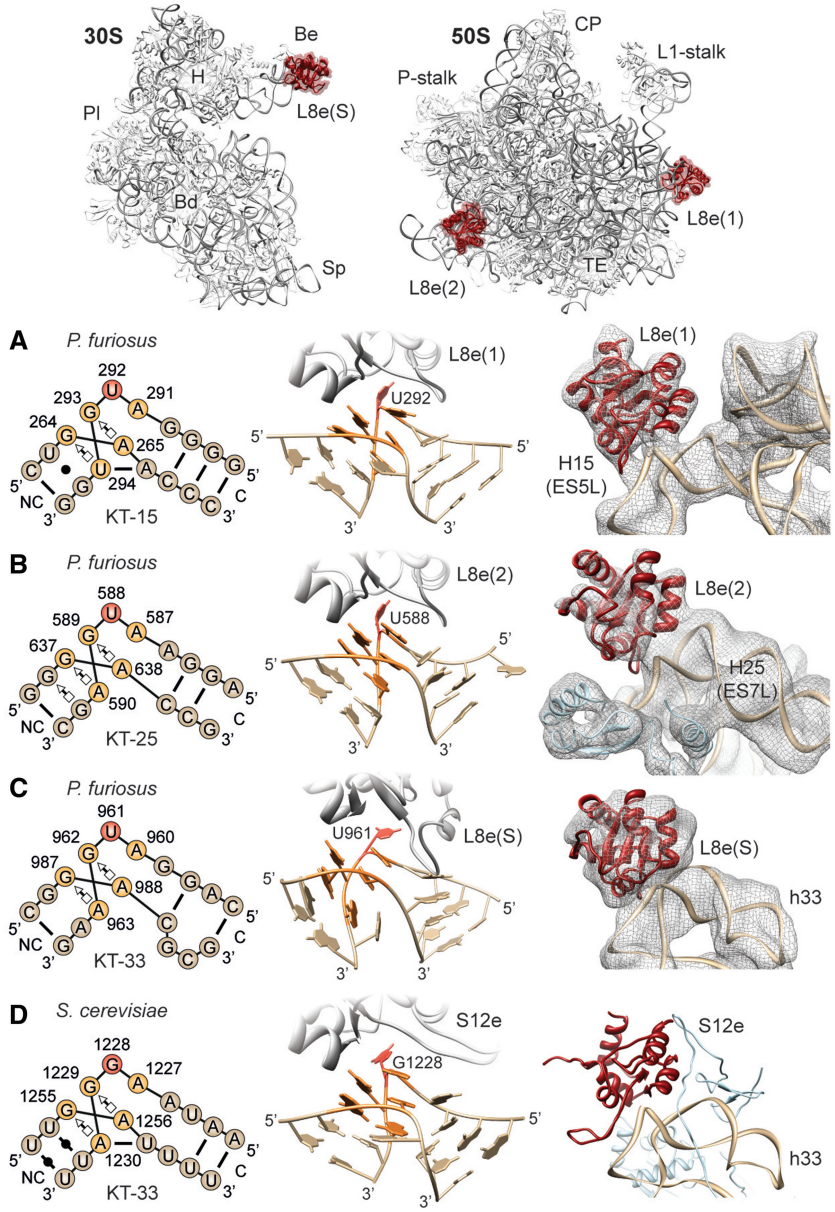


Figure 3. Promiscuity of archaeal ribosomal protein L8e (L7ae). (**A–D**) Schematic (left) and structural (middle) representations of KT motif (A) KT-15 of the canonical L8e(1) binding site and (B) KT-25 at the L8e2 position, on the 50S subunit, compared with KT-33 at the (C) L8e(S) binding site on the *P. furiosus* 30S subunit and (D) S12e binding site on the *S. cerevisiae* 40S subunit (7). Right-hand panels show a fit of molecular models to the cryo-EM density (mesh) of *P. furiosus* 70S ribosome for (A) L8e(1) and (B) L8e(2) on the 50S subunit, and (C) L8e(S) on the small subunit, and in comparison (D), the binding position of S12e on the *S. cerevisiae* 40S subunit (7). Insets at top of figure show the overview of the L8e-binding positions (red) on the small (left) and large (right) ribosomal subunit. Major landmarks are indicated: beak (Be), body (Bd), platform (Pl), head (H), spur (Sp), central protuberance (CP) and tunnel exit (TE). C and NC indicate the canonical and non-canonical stem in KT diagrams. The cross-correlation of the fit of L8e crystal structure to the density for the different binding sites is as follows: L8e(1) = 0.90; L8e(2) = 0.87; L8e(S) = 0.81.

36/38 (95%) of the 30S and 50S subunit r-proteins, respectively (Supplementary Tables S3 and S4). Additionally, the *T. kodakaraensis* 70S ribosomes were split into 30S and 50S subunits, isolated using sucrose gradients and also analyzed by 2D-PAGE and MS (Figure 2C and D). Surprisingly, MS revealed that a protein spot for large subunit r-protein L8e (L7ae) was also present in the 2D-PAGE of the 30S subunit, which, although relatively weak, had similar intensity to some *bona fide* small subunit r-proteins, such as S10, S17 and S19e (Figure 2C). Similarly, MS identified a protein spot for the small subunit r-protein S24e in the 2D-PAGE of the 50S subunit, with comparable intensity to large subunit r-proteins L11, L16 and L35 (Figure 2D).

Identification and localization of promiscuous archaeal r-protein L8e (L7ae)

L8e (L7ae) is a compact globular protein with a fold consisting of alternating α-helices and β-strands, forming an α-β-α sandwich structure that interacts with a kink-turn (KT) motif (39,40)—an asymmetric internal loop that induces a sharp bend in the phoshodiester backbone of an RNA helix (39). In the archaeal and eukaryotic large subunit structures (7,9,14), L8e interacts with the KT motif present in helix 15 of the 23S rRNA (termed KT-15) by specifically recognizing a bulged uridine nucleotide (U292) located within the internal loop (Figure 3A). Thus, to ascertain whether L8e has an additional binding site on the small and/or large subunit, we searched for similar KT motifs within the 16S and 23S rRNAs of the *T. kodakaraensis* and *P. furiosus* 70S ribosome. Two further KT motifs with bulged uridine nucleotides were identified in direct vicinity of the unassigned densities in the *P. furiosus* 70S cryo-EM map, one in H25 of the 23S rRNA (KT-25, Figure 3B) and the other in h33 of the 16S rRNA (KT-33, Figure 3C). Docking of the model for L8e into each of the unassigned densities yielded an excellent fit and maintained canonical interaction with the bulged uridine nucleotide of the respective KT motifs (Figure 3B and C). Together with the MS data, these findings suggest that L8e has three binding sites in the *P. furiosus* 70S ribosome, namely, the canonical site, L8e(1), positioned adjacent to the L1 stalk; a second site, L8e(2), interacting with H25 (ES7L) at the back of the large subunit; as well as a third site, L8e(S), located on the beak of the small subunit. This promiscuity of L8e already has a precedent, as L8e (L7ae) is known to also interact with bulged uridine nucleotides found within KT motifs of the archaeal C/D and H/ACA box archaeal small nucleolar ribonucleoprotein particle (snoRNP) RNAs (Supplementary Figure S8) (41,42) as well as within the KT-containing archaeal RNase P RNA (43).

In eukaryotic 60S subunits, ES7L is elongated and KT-25 is absent, thus the L8e(2) binding site is not present (7,32,44) (Supplementary Figure S9). In contrast, KT-33 is present in all eukaryotic 40S subunits, where it forms part of the binding site for eukaryotic-specific r-protein S12e (Figure 3D)—a protein with the same fold as L8e (Supplementary Figure S10). In fact, S12e actually belongs to the evolutionary conserved L8e (L7ae) family of KT-binding proteins, which also encompasses r-protein L30e as well as the dual spliceosome/snoRNP 15.5kD protein (Snu13p in yeast), RNase P component Rpp38 (Pop3p), the snoRNP protein NHP2 and SBP2, a protein that binds the selenocysteine insertion sequence element RNA (39,45,46). However, S12e recognizes a bulged-out guanine (G1228) nucleotide (rather than a uridine as L8e) within the internal loop of KT-33 (Figure 3D), which is conserved in eukaryotic 18S rRNAs (Supplementary Figure S11). Substitution of uridine to guanine in KT-containing RNAs reduces binding affinity of archaeal L8e by ~100-fold (47), indicating how eukaryotic 80S ribosomes ensure specificity of L8e and S12e binding to KT-15 and KT-33, respectively.

Identification and localization of promiscuous archaeal r-proteins S24e and L14e

Based on the 2D-PAGE and MS data, we examined whether r-protein S24e also has a binding site on the 50S subunit that could account for one of the remaining unassigned densities in the *P. furiosus* 70S cryo-EM map. S24e comprises a four-stranded anti-parallel β-sheet flanked by three short α-helices and contains an RNA

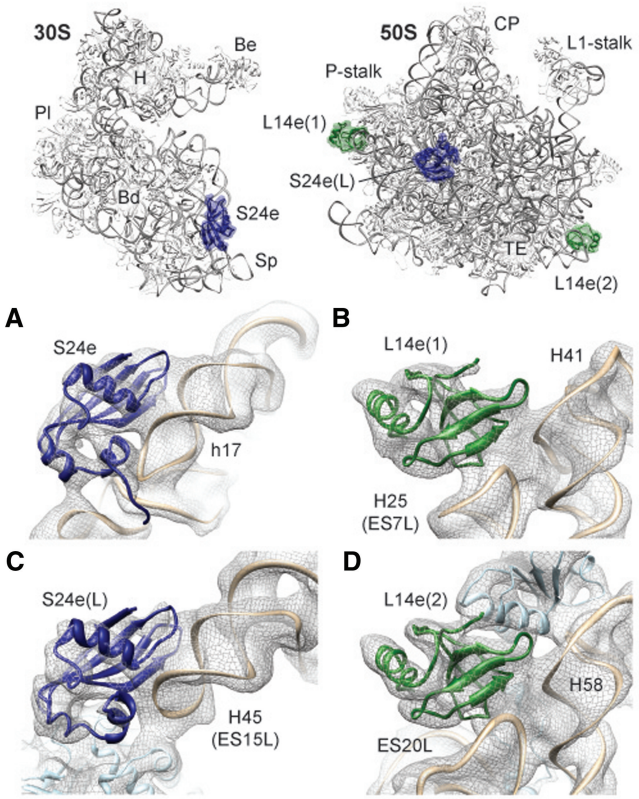


Figure 4. Promiscuity of archaeal ribosomal proteins S24e and L14e. (**A–D**) Fit of molecular models of (A-B) S24e (blue) and (C-D) L14e (green) into the cryo-EM density (gray mesh) of *P. furiosus* 70S ribosome, namely, (A) S24e on the 30S subunit and (B) S24e(S) on the 50S subunit, and (C) L14e(1) and (D) L14e(2) on the 50S subunit. Insets at the top show the overview of the S24e (blue) and L14e (green)-binding positions on the small (left) and large (right) ribosomal subunit. Major landmarks are indicated as in Figure 3.

recognition motif, similar to the related r-protein L23 (48). The canonical binding site for S24e is located at the base of the body of the small subunit, where the terminal β-strand of S24e interacts with the major groove of h17 (Figure 4A). We found an excellent fit of the model for S24e to one of the unassigned densities on the large subunit, similarly oriented such that the terminal β-strand of S24e interacts with major groove of an RNA helix, in this case H45 (ES15L) (Figure 4B). ES15L is slightly longer and adopts a different conformation in eukaryotic 80S ribosomes, consistent with the absence of a second S24e(2)-binding position (7,9,32) (Supplementary Figure S9).

Given the surprising finding that archaeal r-proteins, such as L8e and S24e, were present in more than one copy per *P. furiosus* 70S ribosome, we next examined whether the last unassigned density could be also attributed to another promiscuous r-protein. Indeed, we found that L14e could be unambiguously fit to the remaining electron density located on the 50S subunit (Figure 4C). Archaeal L14e has an Src-homology 3 (SH3)-like β-barrel (49) and in the canonical position on the ribosome, L14e(1), is located at the back of the central protuberance adjacent to LX, where it interacts with the backbone of H41 and the tip of H25 (ES7L) (Figure 4C) (7,9,15,44). Analogously, the second L14e(2) interacts with the backbone of H58 and the tip of ES20L (Figure 4D). In eukaryotic 80S ribosomes, the binding position of L14e(2) is occupied by the related eukaryotic-specific r-protein L27e, which also contains an SH3-like β-barrel (7,9,14,15) (Supplementary Figure S9).

Taxonomic distribution of the promiscuous archaeal ribosomal proteins

The discovery of inter- and intra-subunit promiscuity and multi-copy r-proteins within the *P. furiosus* 70S ribosome raised the question as to whether this represents a general phenomenon occurring in other archaeal species or whether it is specific for the *Thermococcaceae* family. To address this, we searched for the presence of additional binding sites of L8e, L14e and S24e in the available archaeal ribosomal structures, namely the 50S subunit from the Euryarchaeota *H. marismortui* (14) and *M. thermoautotrophicus* (15) as well as the Crenarchaeota *S. acidocaldarius* and *P. aerophilum* (21). In addition, we determined additional cryo-EM structures of the Euryarchaeota *M. igneus* and *T. kodakaraensis* 70S ribosomes at 18 Å and 25 Å (0.5 FSC), respectively, as well as the Crenarchaeota *S. marinus* 50S subunit at 24 Å (0.5 FSC) resolution. With the exception of the cryo-EM structure of the *T. kodakaraensis* 70S ribosome, we did not observe additional density in any of the cryo-EM maps for the presence of S24e on the large subunit (Supplementary Figure S12A–F), suggesting that S24e(L) is specific for the *Thermococcaceae* family. Consistently, S24e(L) was also not observed in the X-ray structure of the *H. marismortui* 50S subunit (14). In contrast, we observed additional density for a second binding position of L14e in all the cryo-EM maps (Supplementary Figure S12G–L), suggesting that L14e(2) is ubiquitous across the archaeal phylogeny. However, L14e(2) was not found in the X-ray structure of the *H. marismortui* 50S subunit (14), consistent with the finding that L14e has been lost in *Halobacteria* (10–12).

The presence of additional binding sites of L8e on archaeal ribosomes correlates perfectly with the expectations based on the presence or absence of the relevant KT motif in h33 (KT-33) and H25 (KT-25) of the 16S and 23S rRNA, respectively. Specifically, additional density was observed for L8e(2) on the 50S subunits of *M. thermoautotrophicus*, *S. acidocaldarius*, *S. marinus* and *T. kodakaraensis* (Figure 5A–D), all of which are predicted to contain KT motifs with a conserved uridine in the internal loop (Supplementary Figure S13), whereas no

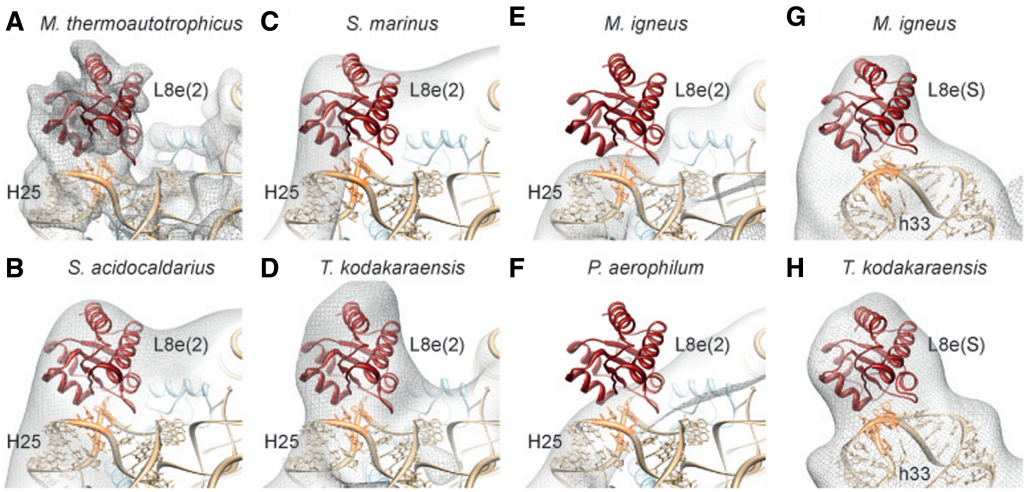


Figure 5. Distribution of additional L8e-binding sites on archaeal ribosomes. (**A–H**) The cryo-EM density (gray mesh) and fitted model for the *P. furiosus* 70S ribosome showing the presence of L8e(2) (red) in (A) *M. thermoautotrophicus* 50S (EMD-2012) (15), (B) *S. acidocaldarius* 50S (EMD-1797) (21), (C) *S. marinus* 50S, (D) 50S subunit of *T. kodakaraensis* 70S and absence in (E) 50S subunit of *M. igneus* 70S, (F) *P. aerophilum* 50S (EMD-1796) (21), as well as presence of L8e(S) in the (G) 30S of *M. igneus* 70S and (H) 30S of *T. kodakaraensis* 70S. Ribosomal RNA is shown in tan with KT motifs in orange.

density was observed for L8e(2) in the cryo-EM maps of *M. igneus* and *P. aerophilum* 50S subunits (Figure 5E–F), where the KT motif has been lost (Supplementary Figure S13). Moreover, the KT motif is also absent in the X-ray structure of the *H. marismortui* 50S subunit, which lacks L8e(2) (14). In contrast, KT-33 with a conserved uridine in the internal loop is predicted for all archaeal 16S rRNA sequences (Supplementary Figure S11). Consistently, we observed additional density for L8e(S) on the small subunit in our two newly determined cryo-EM maps of the Euryarchaeota 70S ribosomes from *M. igneus* and *T. kodakaraensis* (Figure 5G–H). Moreover, previous 2D-PAGE and MS analysis of 30S subunits from the Crenarchaeota *S. acidocaldarius* also detected a protein spot for L8e (L7ae) with a stoichiometry similar to some small subunit r-proteins (21). Collectively, these findings lead us to suggest that L8e(S) is present in all archaeal ribosomes and should be considered as a *bona fide* archaeal small subunit r-protein.

DISCUSSION

With the exception of the stalk proteins (L7/L12 in bacteria and P1-P3 in eukaryotes), r-proteins are thought to be present in one copy per ribosome. Here, we demonstrate that this concept does not hold true for archaeal ribosomes by showing that *P. furiosus* and *T. kodakaraensis* 70S ribosomes have two copies each of S24e and L14e as well as three copies of L8e. Moreover, based on our analysis of KT motifs across the complete archaeal phylogeny (Figure 6A), we predict that all archaeal ribosomes contain at least two copies of L8e, one at the canonical site on the 50S subunit and an additional site located on the small subunit (KT-33), whereas the second L8e binding site on the 50S subunit (KT-25) appears to be lost predominantly in late branching Euryarchaeota, such as *Methanococcus*, *Archaeoglobus* and *Halobacterium*. To our knowledge, the inter-subunit and intra-subunit promiscuity exhibited by S24e/L8e and L14e/L8e, respectively, has not been observed previously (Figure 6B). Furthermore, we find that the intersubunit promiscuity of S24e is specific for the *Thermococcus*/*Pyrococcus* 70S ribosomes, whereas the intrasubunit promiscuity of L14e appears to be widely established in archaea. The observation that L8e(S) and L14e(2) occupy the same binding position on the archaeal ribosome, as the related r-proteins S12e and L27e occupy on the eukaryotic ribosome (Figure 6B and C and Supplementary Figure S9), may reflect that S12e and L27e originally evolved from the promiscuous behavior of L8e and L14e, as has been suggested for archaeal LX and the related eukaryotic r-protein L20e (15) (Figure 6B and C and Supplementary Figure S9). Such a scenario would be consistent with the intermediate complexity of archaeal ribosomes compared with bacterial and eukaryotic ribosomes (Supplementary Figure S14) reflecting the potential for archaeal ribosomes to represent intermediate steps in the evolution of eukaryotic ribosomes.

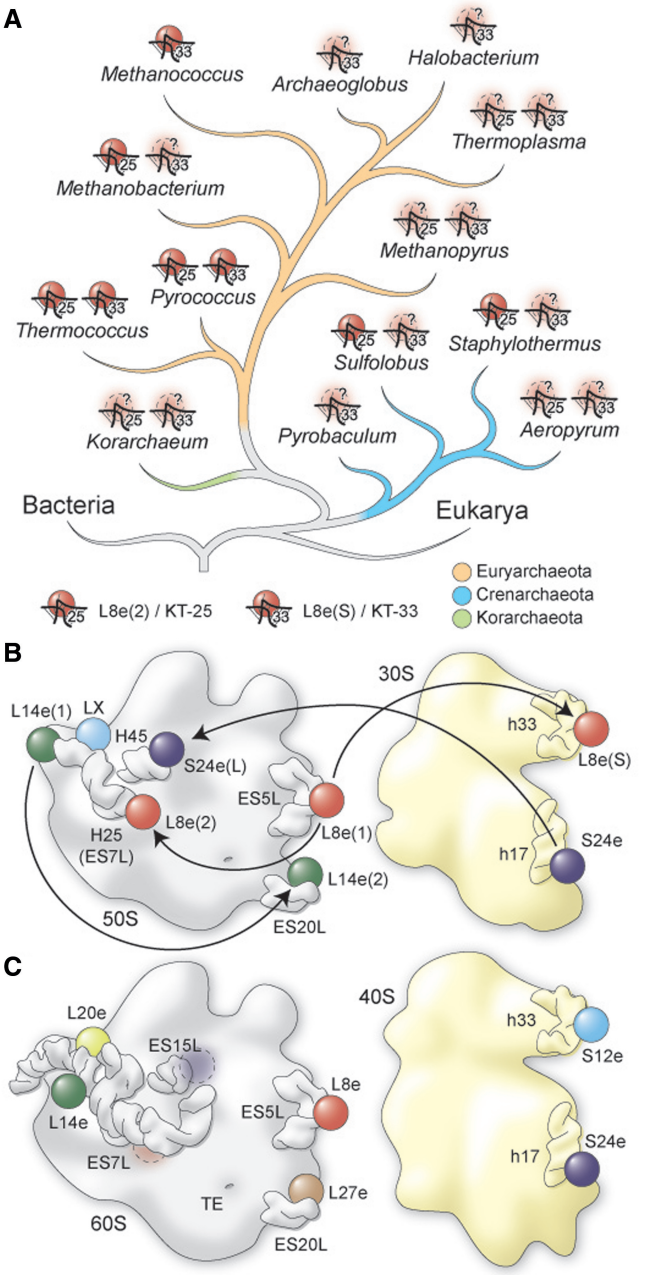


Figure 6. Inter- and intra-subunit promiscuity of archaeal ribosomal proteins. (**A**) Taxonomy of archaea, with presence or absence of KT-25, KT-33 and L8e(2)- or L8e(S)-binding sites. Dashed protein symbols indicate that their presence is predicted, but not experimentally proven yet. Phylogenetic tree is based on (50). (**B**) Schematic showing promiscuity of archaeal ribosomal proteins, compared with (**C**) the equivalent situation in the eukaryotic yeast 80S ribosome.

ACCESSION NUMBERS

The cryo-EM maps of the *M. igneus* 70S, *T. kodakaraensis* 70S and *S. marinus* 50S subunit have been deposited in the EM databank under accession numbers EMD-2172, EMD-2170 and EMD-2171, respectively, and the atomic models for the *P. furiosus* 30S and 50S subunits have been deposited in the protein databank under accession numbers PDB3j20 and 3j21/3j2L, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4, Supplementary Figures 1–15 and Supplementary references [51–53].

FUNDING

Deutsche Forschungsgemeinschaft [SFB594, SFB646 to R.B. and FOR1805 to R.B. and D.N.W.]; Fonds der chemischen Industrie (to S.F.); EMBO young investigator program (to D.N.W.). Funding for open access charge: LMU.

Conflict of interest statement. None declared.

REFERENCES

1. Schmeing,T.M. and Ramakrishnan,V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature*, **461**, 1234–1242.
2. Melnikov,S., Ben-Shem,A., Garreau de Loubresse,N., Jenner,L., Yusupova,G. and Yusupov,M. (2012) One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.*, **19**, 560–567.
3. Wilson,D.N. and Cate,J.H. (2012) The structure and function of the eukaryotic ribosome. *Cold. Spring. Harb. Perspect. Biol.*, **4**, 1–17.
4. Wimberly,B.T., Brodersen,D.E., Clemons,W.M., Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
5. Schuwirth,B., Borovinskaya,M., Hau,C., Zhang,W., Vila-Sanjurjo,A., Holton,J. and Cate,J. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science*, **310**, 827–834.
6. Selmer,M., Dunham,C., Murphy,F.T., Weixlbaumer,A., Petry,S., Kelley,A., Weir,J. and Ramakrishnan,V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
7. Ben-Shem,A., Garreau de Loubresse,N., Melnikov,S., Jenner,L., Yusupova,G. and Yusupov,M. (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. *Science*, **334**, 1524–1529.
8. Rabl,J., Leibundgut,M., Ataide,S.F., Haag,A. and Ban,N. (2011) Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science*, **331**, 730–736.
9. Klinge,S., Voigts-Hoffmann,F., Leibundgut,M., Arpagaus,S. and Ban,N. (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science*, **334**, 941–948.
10. Lecompte,O., Ripp,R., Thierry,J.C., Moras,D. and Poch,O. (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.*, **30**, 5382–5390.
11. Desmond,E., Brochier-Armanet,C., Forterre,P. and Gribaldo,S. (2010) On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. *Res. Microbiol.*, **162**, 53–70.
12. Yutin,N., Puigbo,P., Koonin,E.V. and Wolf,Y.I. (2012) Phylogenomics of prokaryotic ribosomal proteins. *PLoS One*, **7**, e36972.
13. Becker,T., Franckenberg,S., Wickles,S., Shoemaker,C.J., Anger,A.M., Armache,J.P., Sieber,H., Ungewickell,C., Berninghausen,O., Daberkow,I. *et al.* (2012) Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature*, **482**, 501–506.
14. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
15. Greber,B.J., Boehringer,D., Godinic-Mikulcic,V., Crnkovic,A., Ibba,M., Weygand-Durasevic,I. and Ban,N. (2012) Cryo-EM structure of the archaeal 50S ribosomal subunit in complex with initiation factor 6 and implications for ribosome evolution. *J. Mol. Biol.*, **418**, 145–160.
16. Keller,M., Braun,F.J., Dirmeier,R., Hafenbradl,D., Burggraf,S., Rachel,R. and Stetter,K.O. (1995) Thermococcus alcaliphilus sp. nov., a new hyperthermophilic archaeum growing on polysulfide at alkaline pH. *Arch. Microbiol.*, **164**, 390–395.
17. Burggraf,S., Fricke,H., Neuner,A., Kristjansson,J., Rouvier,P., Mandelco,L., Woese,C.R. and Stetter,K.O. (1990) *Methanococcus igneus* sp. nov., a novel hyperthermophilic methanogen from a shallow submarine hydrothermal system. *Syst. Appl. Microbiol.*, **13**, 263–269.
18. Balch,W.E., Fox,G.E., Magrum,L.J., Woese,C.R. and Wolfe,R.S. (1979) Methanogens: reevaluation of a unique biological group. *Microbiol. Rev.*, **43**, 260–296.
19. Londei,P., Altamura,S., Cammarano,P. and Petrucci,L. (1986) Differential features of ribosomes and of poly(U)-programmed cell-free systems derived from sulfur-dependent archaebacterial species. *Eur. J. Biochem.*, **157**, 455–462.
20. Bommer,U., Burkhardt,N., Jünemann,R., Spahn,C.M.T., Triana-Alonso,F.J. and Nierhaus,K.H. (1996) In: Graham,J. and Rickwoods,D. (eds), *Subcellular Fractionation. A Practical Approach*. IRL Press at Oxford University Press, Oxford, pp. 271–301.
21. Marquez,V., Frohlich,T., Armache,J.P., Sohmen,D., Donhofer,A., Mikolajka,A., Berninghausen,O., Thomm,M., Beckmann,R., Arnold,G.J. *et al.* (2011) Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J. Mol. Biol.*, **405**, 1215–1232.
22. Nierhaus,K.H. and Dohme,F. (1974) Total reconstitution of functionally active 50S ribosomal subunits from *E. coli*. *Proc. Natl Acad. Sci. USA*, **71**, 4713–4717.
23. Kaltschmidt,E. and Wittmann,H.G. (1970) Ribosomal proteins. VII 2D polyacrylamide gel electrophoresis for fingerprinting of ribosomal proteins. *Anal. Biochem.*, **36**, 401–412.
24. Wagenknecht,T., Frank,J., Boublik,M., Nurse,K. and Ofengand,J. (1988) Direct localization of the tRNA-anticodon interaction site on the *Escherichia coli* 30 S ribosomal subunit by electron microscopy and computerized image averaging. *J. Mol. Biol.*, **203**, 753–760.
25. Mindell,J.A. and Grigorieff,N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.*, **142**, 334–347.
26. Frank,J., Radermacher,M., Penczek,P., Zhu,J., Li,Y., Ladjadj,M. and Leith,A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.*, **116**, 190–199.
27. Jossinet,F. and Westhof,E. (2005) Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics*, **21**, 3320–3321.
28. Brodersen,D.E., Clemons,W.M. Jr, Carter,A.P., Wimberly,B.T. and Ramakrishnan,V. (2002) Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.*, **316**, 725–768.
29. Jin,H., Kelley,A.C., Loakes,D. and Ramakrishnan,V. (2010) Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. *Proc. Natl Acad. Sci. USA*, **107**, 8593–8598.
30. Dunkle,J.A., Wang,L., Feldman,M.B., Pulk,A., Chen,V.B., Kapral,G.J., Noeske,J., Richardson,J.S., Blanchard,S.C. and Cate,J.H. (2011) Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science*, **332**, 981–984.
31. Jossinet,F., Ludwig,T.E. and Westhof,E. (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, **26**, 2057–2059.
32. Armache,J.P., Jarasch,A., Anger,A.M., Villa,E., Becker,T., Bhushan,S., Jossinet,F., Habeck,M., Dindar,G., Franckenberg,S. *et al.* (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. *Proc. Natl Acad. Sci. USA*, **107**, 19748–19753.
33. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

34. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
35. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
36. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
37. Trabuco,L.G., Villa,E., Mitra,K., Frank,J. and Schulten,K. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, **16**, 673–683.
38. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD—Visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.
39. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
40. Klein,D., Moore,P. and Steitz,T. (2004) The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.*, **340**, 141–177.
41. Moore,T., Zhang,Y., Fenley,M.O. and Li,H. (2004) Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure*, **12**, 807–818.
42. Li,L. and Ye,K. (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, **443**, 302–307.
43. Cho,I.M., Lai,L.B., Susanti,D., Mukhopadhyay,B. and Gopalan,V. (2010) Ribosomal protein L7Ae is a subunit of archaeal RNase P. *Proc. Natl Acad. Sci. USA*, **107**, 14573–14578.
44. Armache,J.P., Jarasch,A., Anger,A.M., Villa,E., Becker,T., Bhushan,S., Jossinet,F., Habeck,M., Dindar,G., Franckenberg,S. *et al.* (2010) Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. *Proc. Natl Acad. Sci. USA*, **107**, 19754–19759.
45. Dennis,P.P. and Omer,A. (2005) Small non-coding RNAs in Archaea. *Curr. Opin. Microbiol.*, **8**, 685–694.
46. Dlakic,M. (2005) 3D models of yeast RNase P/MRP proteins Rpp1p and Pop3p. *RNA*, **11**, 123–127.
47. Kuhn,J.F., Tran,E.J. and Maxwell,E.S. (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.*, **30**, 931–941.
48. Choesmel,V., Fribourg,S., Aguissa-Toure,A.H., Pinaud,N., Legrand,P., Gazda,H.T. and Gleizes,P.E. (2008) Mutation of ribosomal protein RPS24 in Diamond-Blackfan anemia results in a ribosome biogenesis disorder. *Hum. Mol. Genet.*, **17**, 1253–1263.
49. Edmondson,S.P., Turri,J., Smith,K., Clark,A. and Shriver,J.W. (2009) Structure, stability, and flexibility of ribosomal protein L14e from *Sulfolobus solfataricus*. *Biochemistry*, **48**, 5553–5562.
50. Brochier-Armanet,C., Forterre,P. and Gribaldo,S. (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.*, **14**, 274–281.
51. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Muller,K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2.
52. Xue,S., Wang,R., Yang,F., Terns,R.M., Terns,M.P., Zhang,X., Maxwell,E.S. and Li,H. (2010) Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol. Cell*, **39**, 939–949.
53. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.

Supplementary Information

Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution

Jean-Paul Armache[1,2*], Andreas M. Anger[1,2*], Viter Márquez[1,2#], Sibylle Franckenberg[1,2#], Thomas Fröhlich[1,3], Elizabeth Villa[4], Otto Berninghausen[1,2], Michael Thomm[5], Georg J. Arnold[1,3], Roland Beckmann[1,2,6] and Daniel N. Wilson[1,2,6]

[1] Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor Lynen Str. 25, 81377, Munich, Germany

[2] Center for integrated Protein Science Munich (CiPSM), Ludwig-Maximilians-Universität München, Germany

[3] Laboratory for Functional Genome Analysis LAFUGA, Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor Lynen Str. 25, 81377, Munich, Germany

[4] Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

[5] Lehrstuhl für Mikrobiologie, Regensburg University, Universitätstraße 32, 93053 Regensburg, Germany

[6] Corresponding authors: wilson@lmb.uni-muenchen.de, beckmann@lmb.uni-muenchen.de

* These authors contributed equally to this work

# These authors contributed equally to this work

Supplementary Table 1. Distribution of ribosomal proteins on the ribosomal small subunit in Eukarya (E), Korarchaeota (KA), Crenarchaeota (CA), Euryarchaeota (EA) and Bacteria (B)

Protein names	E		KA	CA				EA								B	
	Sce	*Hsa*	*Kcr*	*Ape*	*Sma*	*Eco*	*Pae*	*Pfu*	*Tko*	*Mig*	*Mth*	*Mka*	*Afu*	*Hma*	*Tvo*	*Tth*	*Eco*
RACK1	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S1e (S3ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S4	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S4e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S6e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S7	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S7e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S8	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S8e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S9	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S10	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S10e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S11	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S12	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S12e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S13	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S14	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S15	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S17	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S17e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S19	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S19e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S21e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S24e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S25e	√	√	–	√	√	√	√	–	–	–	–	–	–	–	–	–	–
S26e	√	√	√	√	√	√	√	–	–	–	–	–	–	–	–	–	–
S27e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S28e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–	–
S30e	√	√	√	√	√	√	√	–	–	–	–	–	–	–	–	–	–
S31e (S27ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	Thx	–
L8eS (L7ae)	–	–	u	u	√	√	√	√	√	√	√	u	u	u	√	–	–

Hsa (Homo sapiens), Kcr (Korarchaeum cryptofilum, Ape (Aeropyrum pernix), Sma (Staphylothermus marinus), Sac (Sulfolobus acidocaldarius), Pae (Pyrobaculum aerophilum), Pfu (Pyrococcus furiosus), Tko (Thermococcus kodakaraensis), Mig (Methanococcus igneus), Mth (Methanobacterium thermautotrophicus), Mka (Methanopyrus kandleri), Afu (Archaeoglobus fulgidus), Tvo (Thermoplasma volcanii), Hma (Haloarcula marismortui), Eco (Escherichia coli)

Thx Protein known to exist in Tth belonging to bacterial S31e family, but unrelated to eukaryotic S31e

√ Present

– Absent

u Distribution unknown

Supplementary Table 2. Distribution of large subunit r-proteins in Eukarya (E), Korarchaeota (KA), Crenarchaeota (CA), Euryarchaeota (EA) and Bacteria (B)

Protein names	E		KA	CA				EA								B	
	Sce	*Hsa*	*Kcr*	*Ape*	*Sma*	*Sac*	*Pae*	*Pfu*	*Tko*	*Mig*	*Mth*	*Mka*	*Afu*	*Hma*	*Tvo*	*Eco*	*Tth*
L1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L4	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L6	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L6e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
L8e(L7ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L11	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L13	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L13e	√	√	√	√	√	– *	√	–	–	–	–	–	–	–	–	–	–
L14	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L14e	√	√	√	√	√	√	√	√	√	√	√	√	–	–	–	–	–
L15	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L15e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L16	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L18	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L18e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L19e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L20e(L18ae)	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
L21e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L22	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L22e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
L23	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L24	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L24e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L27e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
L28e	–	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
L29	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L29e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
L30	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L30e	√	√	√	√	√	√	√	√	√	√	√	–	√	–	–	–	–
L31e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L32e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L33e(L35ae)	√	√	–	√	√	–	–	√	√	–	–	–	–	–	–	–	–
L34e	√	√	√	√	√	√	√	√	√	√	√	√	–	–	–	–	–
L36e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
L37e	√	√	√	√	√	√	–	√	√	√	√	√	√	√	√	–	–
L38e	√	√	–	√	√	–	√	–	–	–	–	–	–	–	–	–	–
L39e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L40e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–	–
L41e	√	√	–	–	–	–	–	√	√	√	–	–	–	–	–	–	–
L43e(L37ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L44e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
P0(L10)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
P1(L12)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
P2(L12)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
L8e2(L7ae)	–	–	u	u	√	√	–	√	√	–	√	u	u	– **	u	–	–
L14e2(L14e)	–	–	u	u	√	√	√	√	√	√	√	u	–	–	–	–	–
LX(L20e)	–	–	–	√	√	√	√	√	√	√	√	√	√	√	–	–	–
S24eL	–	–	u	u	–	–	–	√	√	–	–	u	u	–	u	–	–

* Present in *Sulfolobus* species, but not found *S. acidocaldarius*, ** Unknown whether it is absent, or less stably bound, √ Present, − Absent, u Distribution unknown.

Supplementary Table 3. *T. kodakaraensis* 30S r-proteins identified by LC-MS and 2D-PAGE

Protein Name (Family)	LC/MS	2D	Amino acids	gi[a]	Molecular mass (kDa)	pI
S1e (S3ae)	√		200	57641189	23.02	10.1
S2	√	√	201	57641431	23.01	8.9
S3	√	√	209	57641471	23.4	9.6
S4	√	√	180	57641440	21.2	10.2
S4e	√	√	243	57641464	27.8	9.9
S5	√	√	235	57641456	26.3	9.4
S6e	√	√	125	57641886	13.7	10.1
S7	√	√	215	57641012	24.5	10.0
S8	√	√	130	57641461	14.6	9.5
S8e	√	√	130	57641126	14.5	10.8
S9	√	√	135	57641435	15.3	10.5
S10	√	√	102	57640242	11.7	10.1
S11	√	√	140	57641439	15.1	10.3
S12	√	√	147	57641013	16.4	10.6
S13	√	√	149	57641441	16.9	10.8
S14[b]	√		56	57641462	6.6	10.4
S15	√	√	151	57641186	17.5	10.5
S17	√	√	114	57641467	13.2	9.5
S17e	√	√	67	57642227	8	10.3
S19	√	√	133	57641473	15.4	10.3
S19e	√	√	150	57641211	17.3	9.4
S24e	√	√	98	57641631	11.4	5.5
S27e	√	√	65	57641034	7.1	9.2
S28e	√	√	70	57641245	7.9	11.0
S31e (S27ae)	√		57	57641630	6.7	10.2

[a] gi refers to the GenInfo identifier for retrieval from NCBI

[b] Identified with only a single peptide

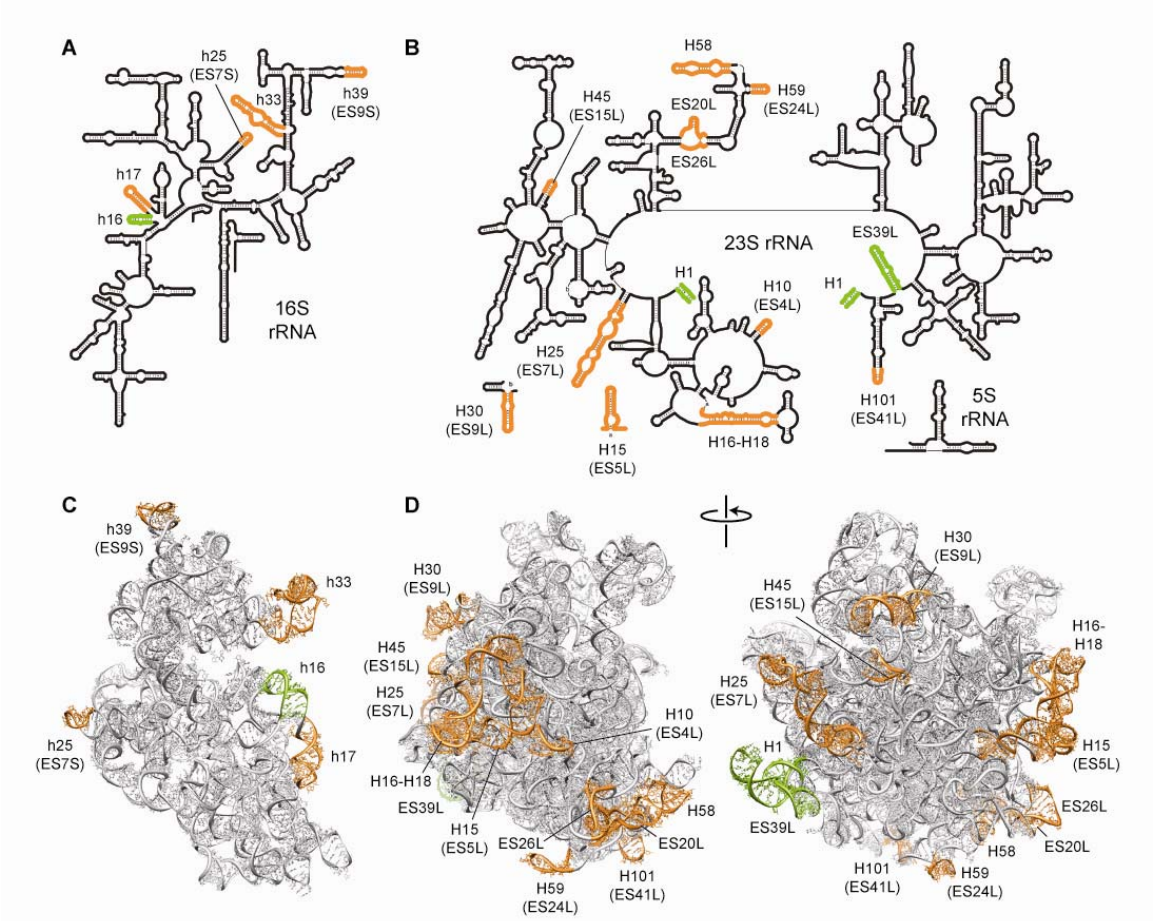
Supplementary Table 4. *T. kodakaraensis* 50S r-proteins identified by LC-MS and 2D-PAGE

Protein Name (Family)	LC/MS	2D	Amino acids	gi[a]	Molecular mass (kDa)	pI
L1	√	√	216	57641352	24.1	9.7
L2	√	√	239	57641474	26.0	10.7
L3	√	√	346	57641477	39	10
L4	√	√	255	57641476	28.7	10.5
L5	√	√	183	57641463	20.9	9.8
L6	√	√	184	57641460	20.8	9.2
L8e (L7ae)	√	√	125	57641246	13.7	5.2
L11	√	√	165	57641353	17.6	5.4
L12			106	57641350	10.8	3.9
L13	√	√	142	57641436	16.3	10.3
L14	√	√	141	57641466	15.2	11.5
L14e	√	√	83	57641448	8.9	9.9
L15	√	√	148	57641454	16.5	10.2
L15e	√	√	194	57641389	22.6	11.2
L16 (L10e)	√	√	182	57641481	21.1	10.3
L18	√	√	201	57641457	22.9	6.5
L18e	√	√	121	57641437	13.8	10.4
L19e	√	√	150	57641458	17.6	10.9
L20e (LX)	√		77	57641257	9.2	9.0
L21e	√	√	98	57640837	11.3	11.5
L22	√	√	156	57641472	17.8	10.7
L23	√	√	86	57641475	9.9	9.7
L24	√	√	121	57641465	14.2	10.0
L24e	√	√	67	57641244	8.1	10.3
L29	√	√	66	57641470	7.9	10.3
L30	√	√	155	57641455	17.8	10.2
L30e[b]	√	√	102	57641015	10.9	8.9
L31e	√	√	90	57641255	10.3	10.6
L32e	√	√	126	57641459	14.7	11.5
L34e	√		90	57641450	10.5	11.7
L33e (L35ae)	√	√	86	57640904	9.5	10.9
L43e (L37ae)	√	√	86	57640550	9.2	11.6
L37e	√		63	57640910	7.5	11.8
L39e	√		51	57641254	6.2	12.6
L40e	√	√	51	57641430	5.8	10.9
L41e			37	57641850	5.0	12.7
L44e	√		94	57641033	11.1	11.1
P0 (L10p)	√		340	57159675	36.8	4.8

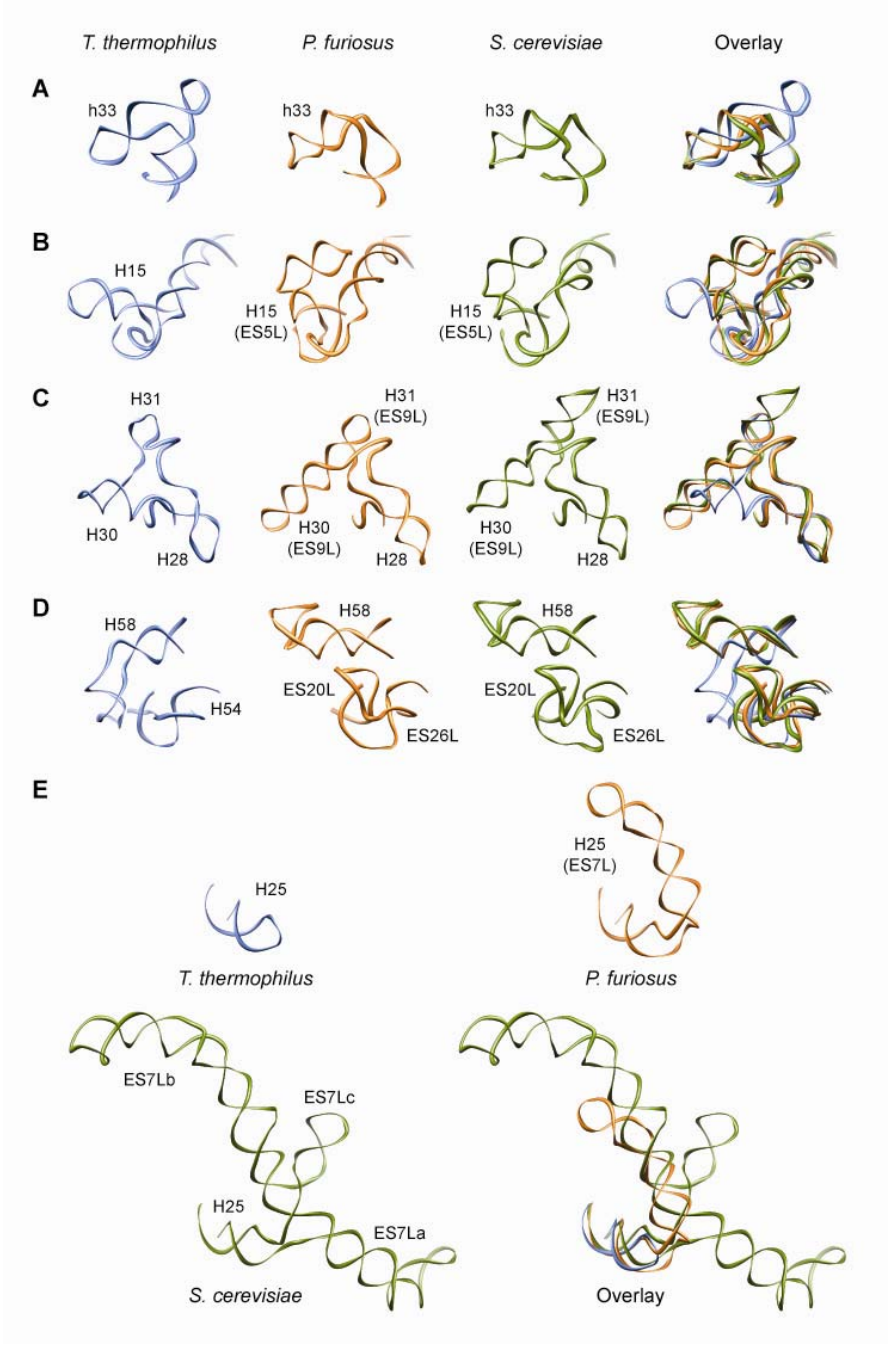
[a] gi refers to the GenInfo identifier for retrieval from NCBI

[b] Identified with only a single peptide

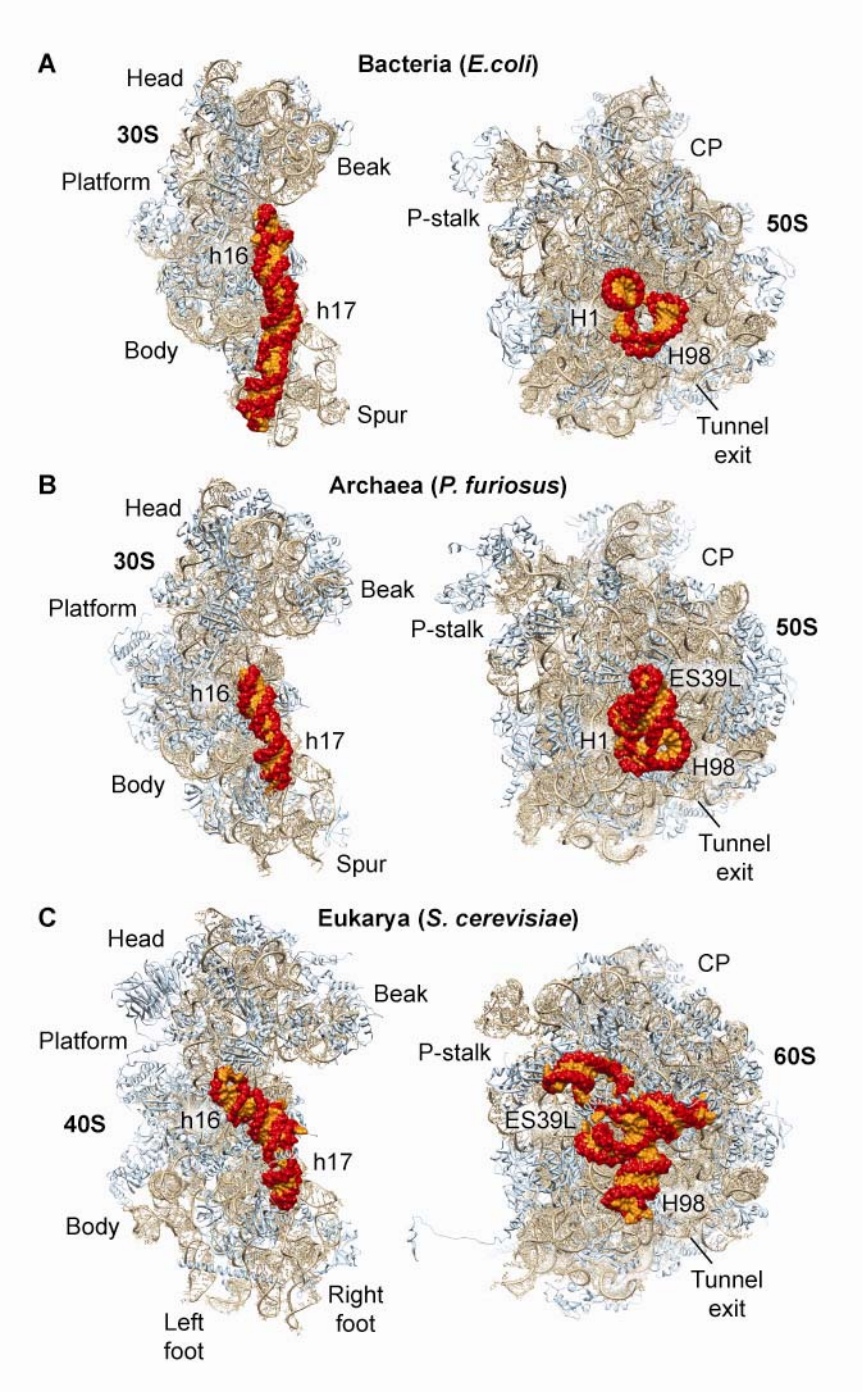
Supplementary Figures



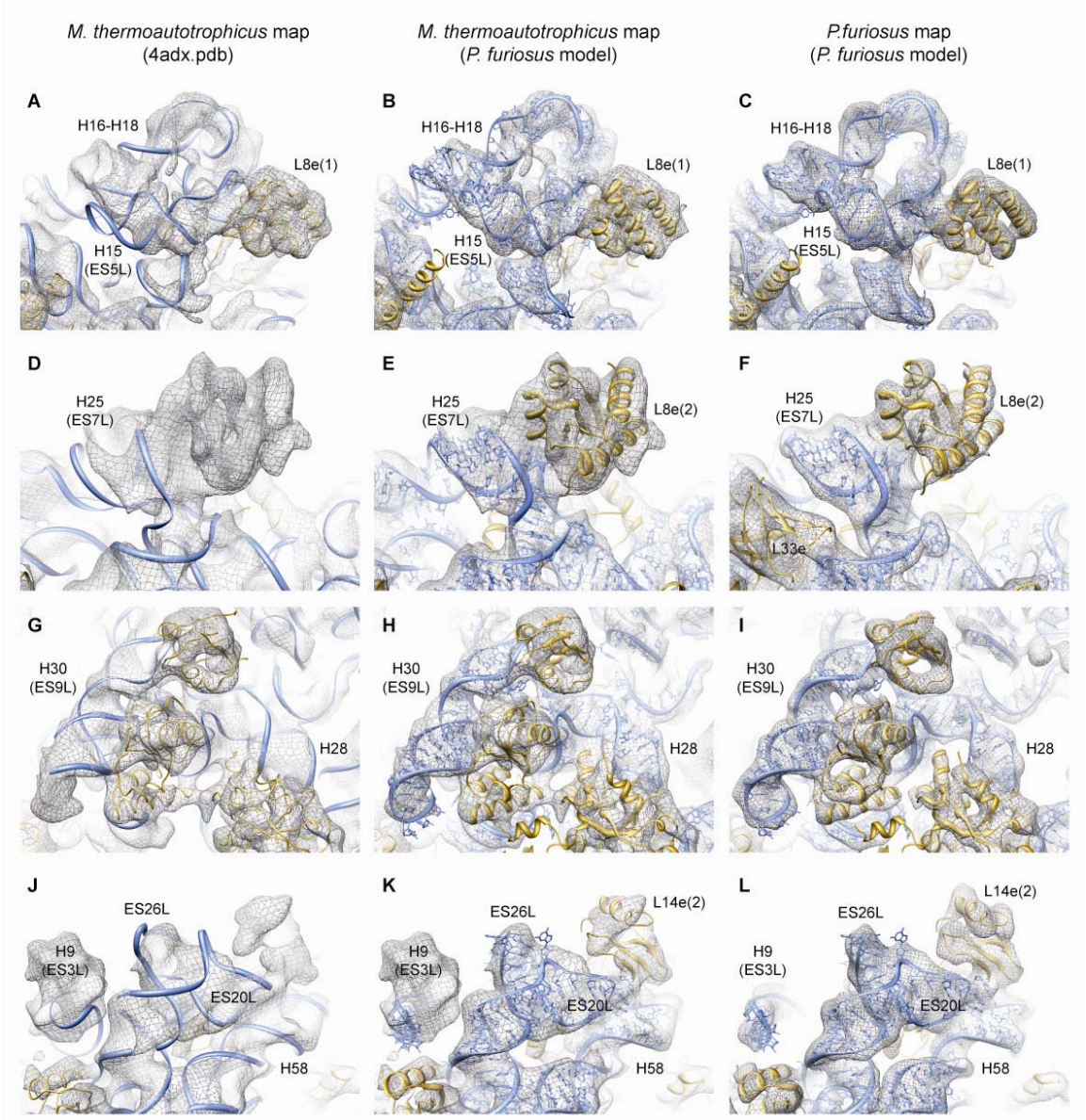
Supplementary Figure 1 Molecular model for rRNA of the *P. furiosus* 70S. (**A-B**) Secondary structure diagrams of the (**A**) 16S and (**B**) 5S and 23S rRNA for *P. furiosus*. Distant parts of the secondary structure drawing are connected by thin lines. (**C-D**) Molecular model of the rRNA for the (**A**) small and (**B**) large subunit of the *P. furiosus* 70S ribosome. Conserved rRNA core is coloured black, eukaryotic-like expansion segments (ES) and variable regions (VR) are coloured orange. ESs and VRs that adopt a unique archaea-specific structure are coloured green.



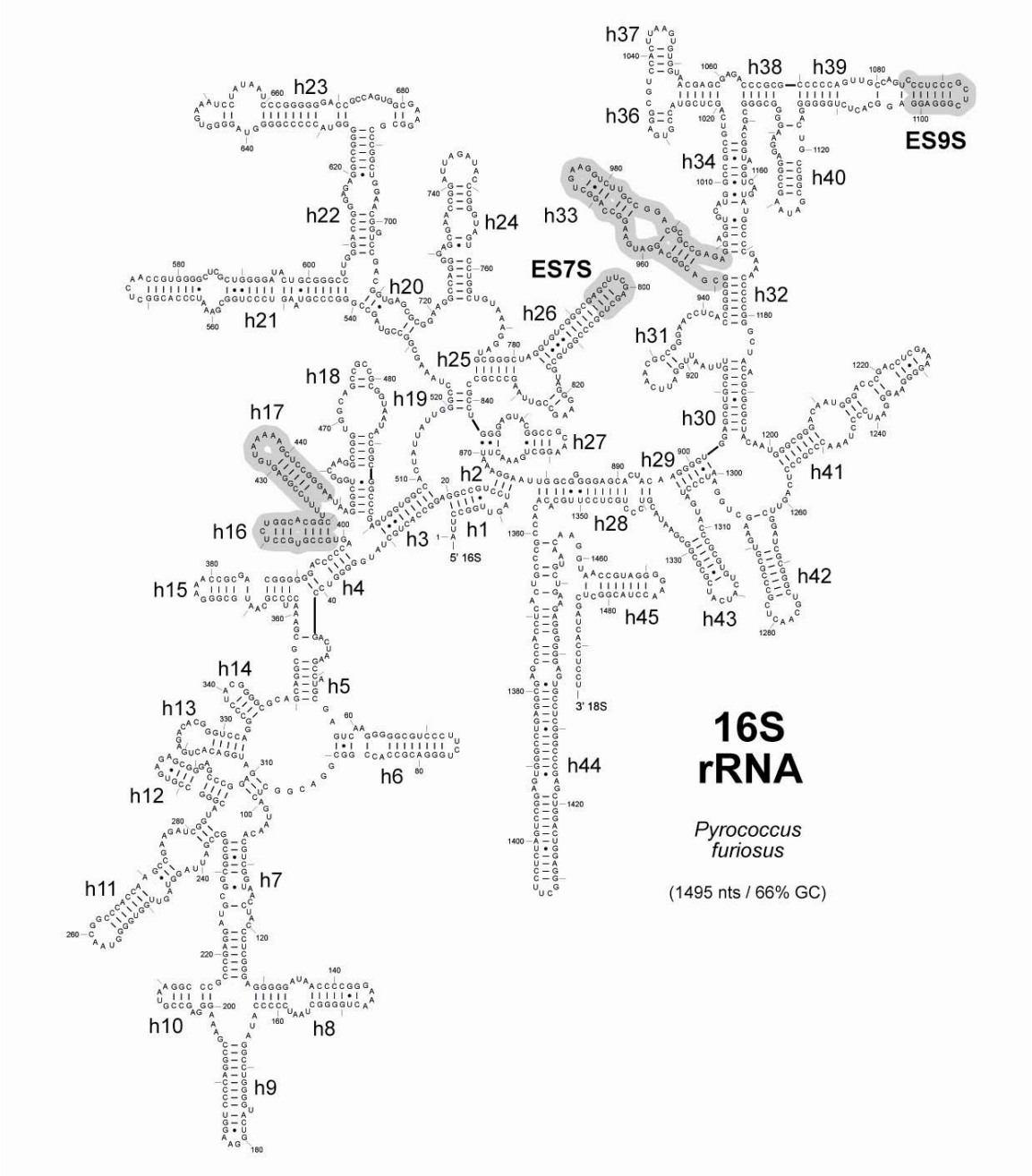
Supplementary Figure 2 Comparison of *P. furiosus* ESs and VRs with bacterial and eukaryotic homologous structures. (**A-E**) Comparison of (A) h33, (B) H15 (ES5L), (C) H28-H31 (ES9L), (D) H54 (ES20L, ES26L), H58 and (E) H25 (ES7L) between the bacteria *T. thermophilus* (blue) (29), the archaea *P. furiosus* (orange), the eukaryote *S. cerevisiae* (green) (7), *S. cerevisiae* ES7L was taken from (32).



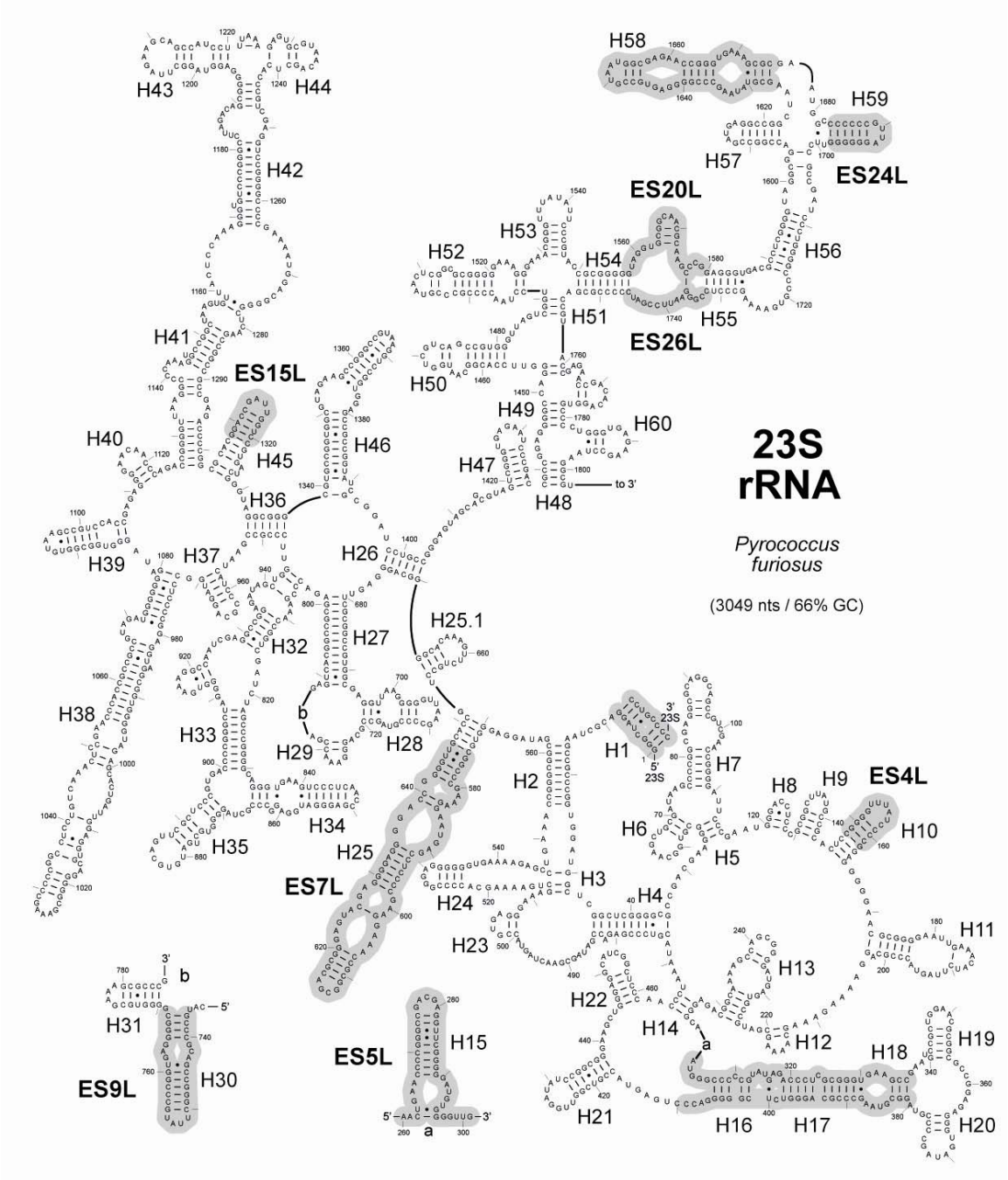
Supplementary Figure 3 Novel rRNA conformations in *P. furiosus* 70S ribosome. (**A-C**) Comparison of h16/17 on the small subunit (left) and H98 (ES39L) on the large subunit (right) between (A) the bacterial (*E. coli*) (30) and (B) archaeal (*P. furiosus*) 70S ribosome, with the (C) eukaryotic (*S. cerevisiae*) 80S ribosome (7). CP marks the central protuberance.



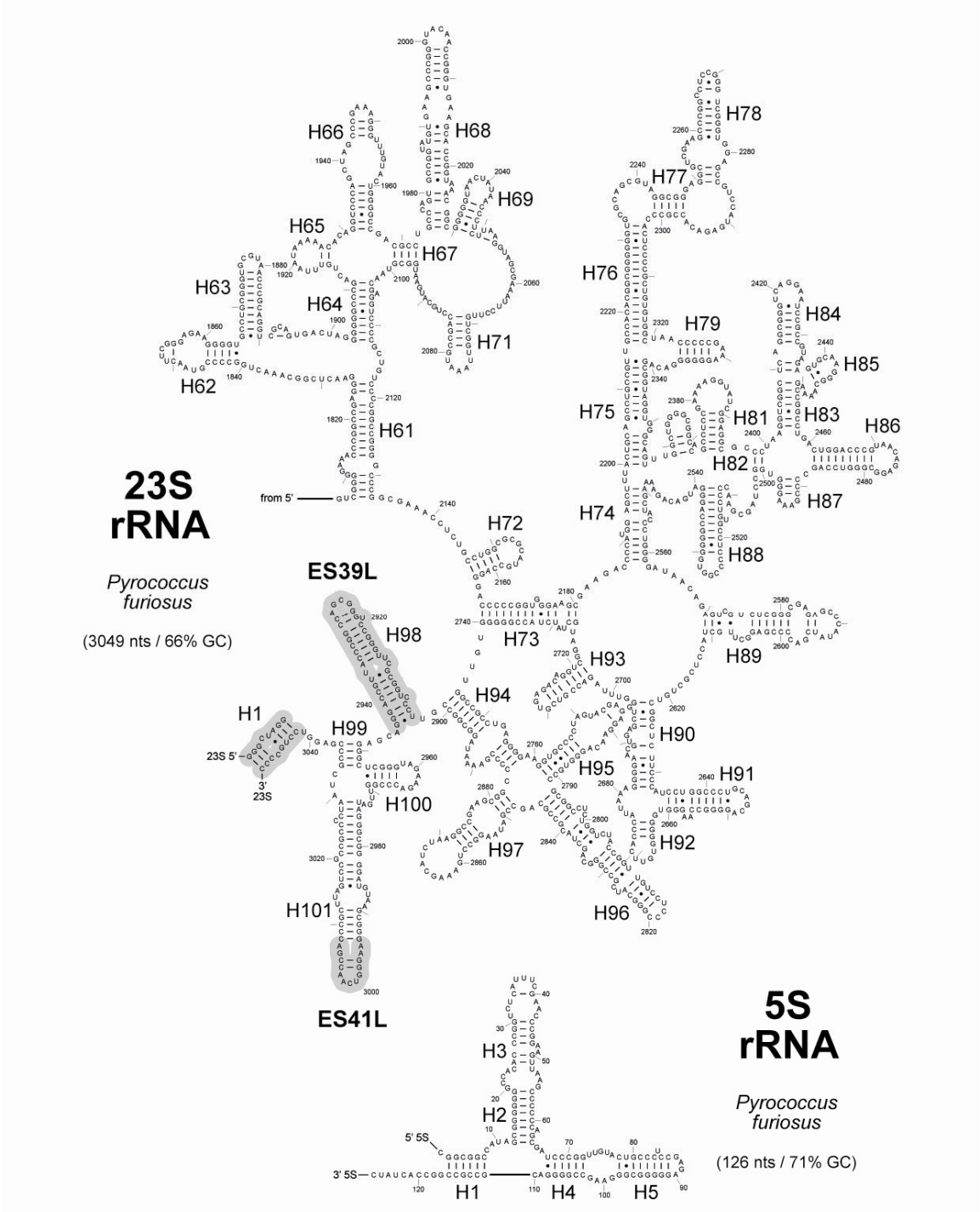
Supplementary Figure 4 Comparison of cryo-EM maps and models for the *M. thermautotrophicus* 50S subunit and the *P. furiosus* 70S ribosome. (**A-L**) Comparison of four distinct regions of the large subunit, namely focusing on (**A-C**) H15 (ES5L), H16-H18 and L8e(1), (**D-F**) H25 (ES7L) and L8e(2), (**G-I**) H28 and ES9L and (**J-L**) ES20L/ES26L and L14e(2). Left panel (A,D,G,J) presents *M. thermautotrophicus* map (EMD-2012) and associated model (PDB4 ADX) (15), middle panel (B,E,H,K) displays *M. thermautotrophicus* map with *P. furiosus* model, and right panel (C,F,I,L) presents the *P. furiosus* map and model. In each case, the cryo-EM density (grey mesh) is shown with rRNA (blue) and proteins (gold).



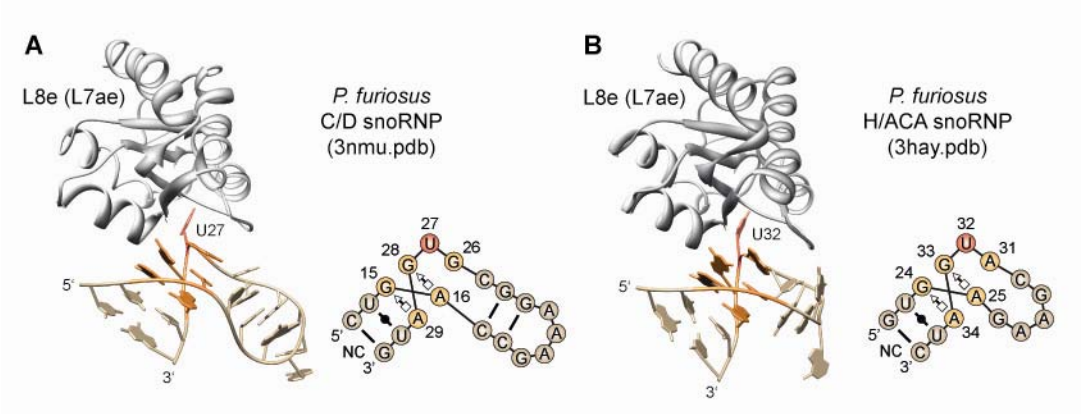
Supplementary Figure 5 Secondary structure diagram of the *P. furiosus* 16S rRNA. The *P. furiosus* 16S rRNA diagram was taken from (51) and adjusted according to the final rRNA model. Grey shaded regions highlight relevant ESs and VRs.



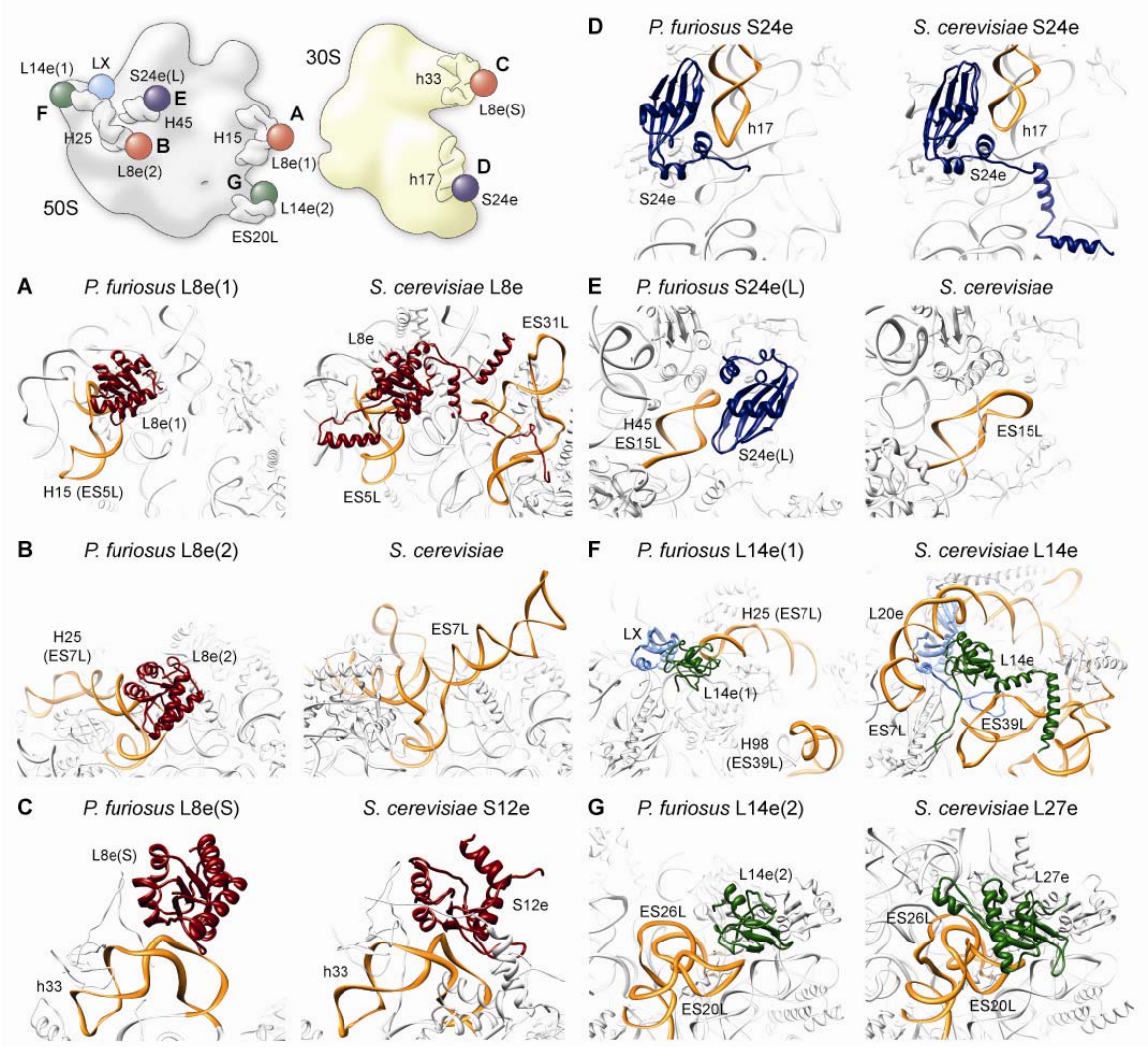
Supplementary Figure 6 Secondary structure diagram of the 5' region of *P. furiosus* 23S rRNA. The *P. furiosus* 23S secondary structure representation is based on the corresponding diagram for *Thermococcus celer* obtained from (51) and was adjusted according to the final rRNA model. Grey shaded regions highlight relevant ESs and VRs.



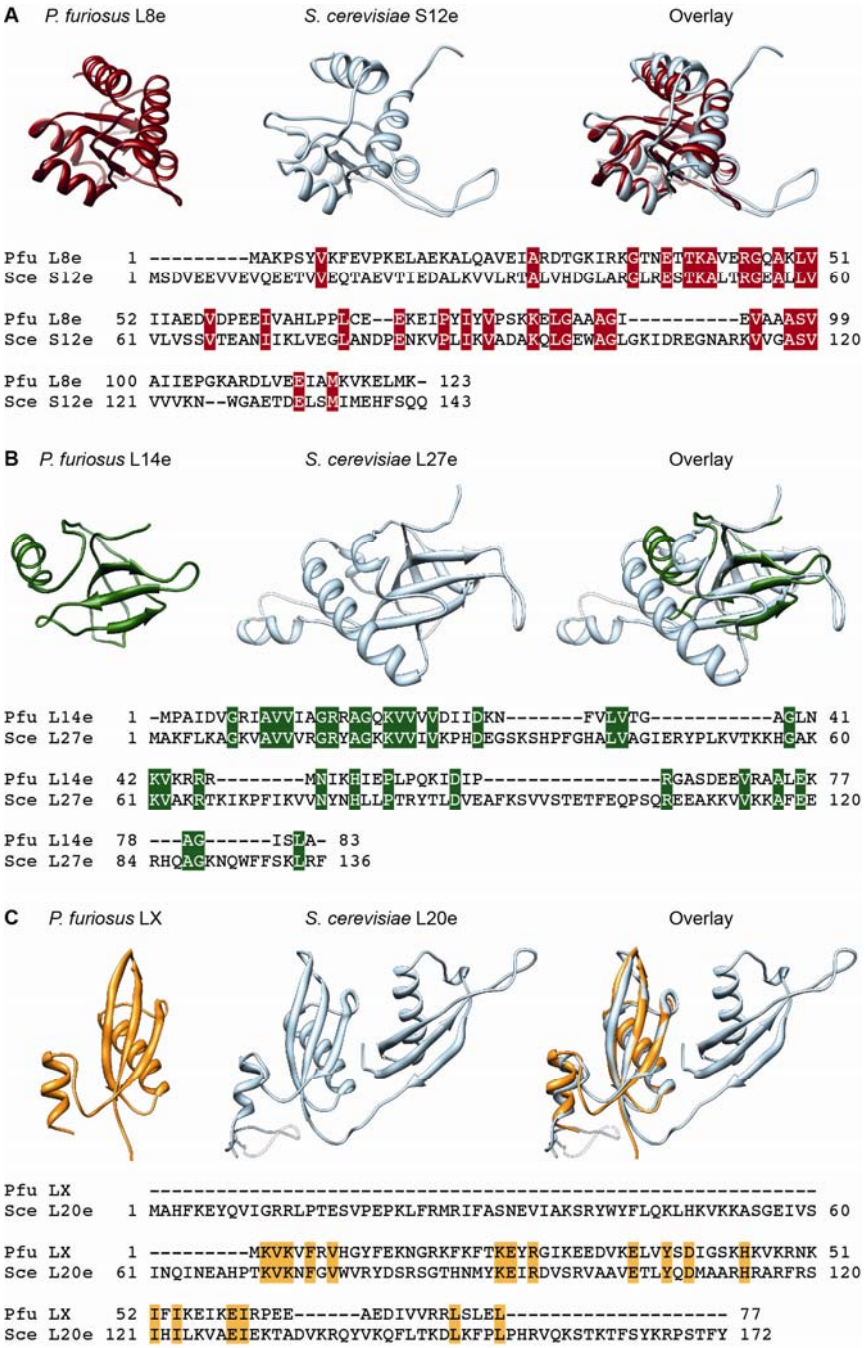
Supplementary Figure 7 Secondary structure diagram of the 3' region of *P. furiosus* 23S rRNA and complete 5S rRNA. The *P. furiosus* 23S and 5S secondary structure representations are based on the corresponding diagrams for *T. celer* and *Pyrococcus woesei*, respectively obtained from (51) and were adjusted according to the final rRNA model. Grey shaded regions highlight relevant ESs and VRs.



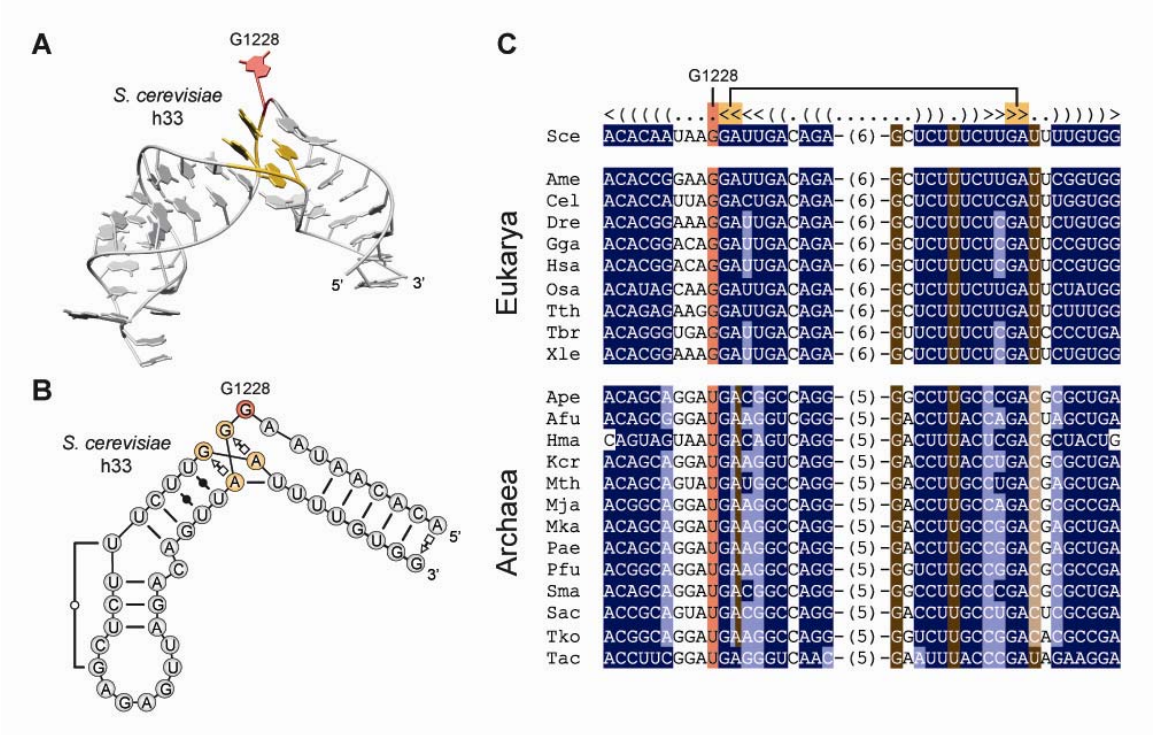
Supplementary Figure 8 *P. furiosus* L8e (L7ae) interaction with kink-turn motifs in C/D and H/ACA snoRNPs. (**A-B**) Interaction of L8e (grey) with bulged out uridine nucleotide (red) within the internal loop of kink-turn motives found in (A) C/D snoRNP (PDB 3NMU) (52) and (B) H/ACA snoRNPs (PDB 2HVY) (42). Insets show secondary structure diagrams of the respective C/D and H/ACA snoRNP kink-turn motives. NC indicates the non-canonical stem.



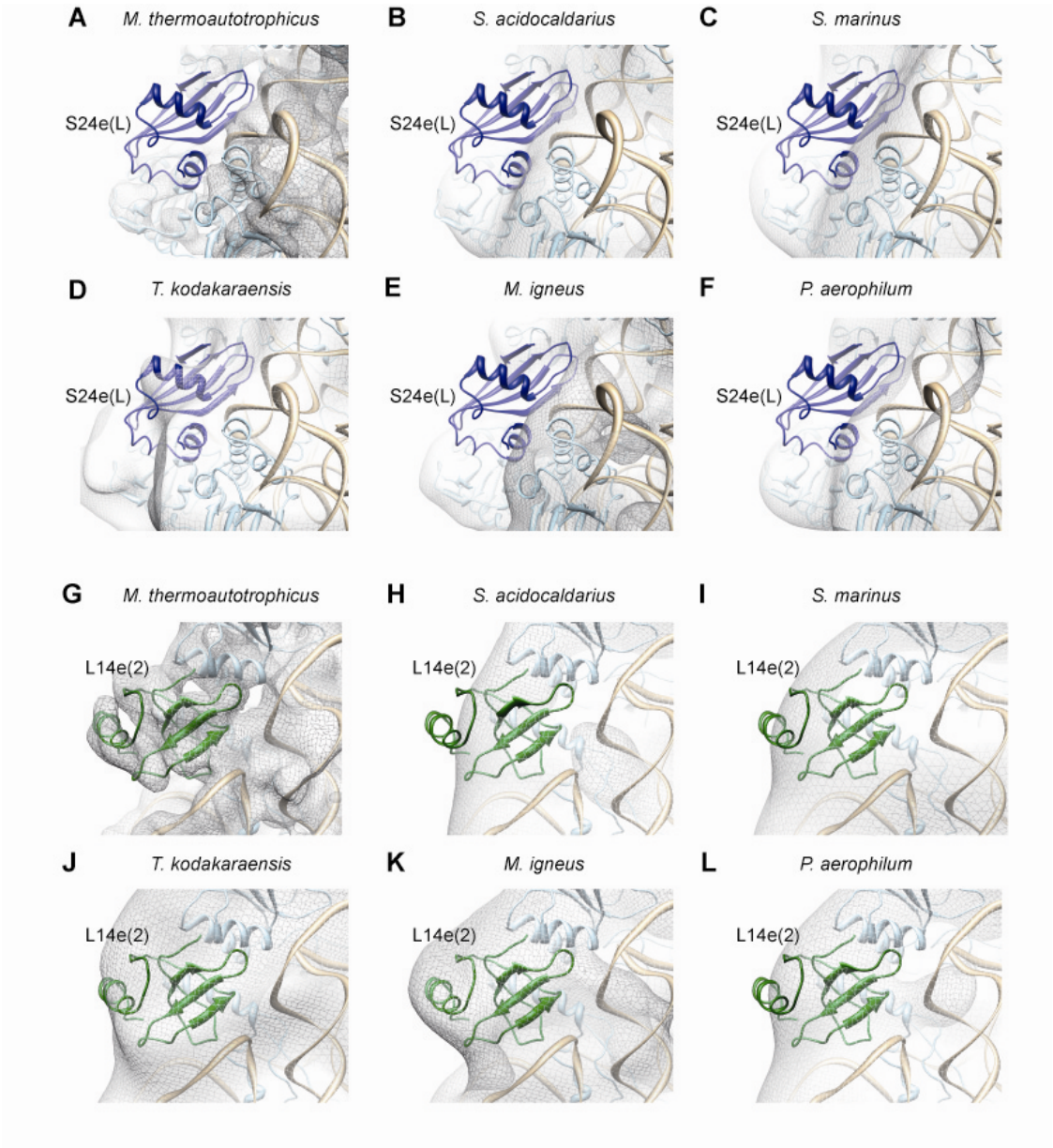
Supplementary Figure 9 Binding positions of *P. furiosus* L8e, L14e and S24e and respective regions on the eukaryotic ribosome. Comparison of binding positions of (A-C) L8e (red), (D-E) S24e (blue) and (F-G) L14e (green) between *P. furiosus* 70S (left panels) and *S. cerevisiae* 80S (right panels) (7) (Note: ES7L was taken from (32)) ribosomes. Relevant rRNA sections shown as orange ribbon, all other ribosomal components are coloured white. Inset shows a schematic view of the location of L8e (red), L14e (green), LX (light blue) and S24e (blue) on the small (yellow) and large (grey) subunit.



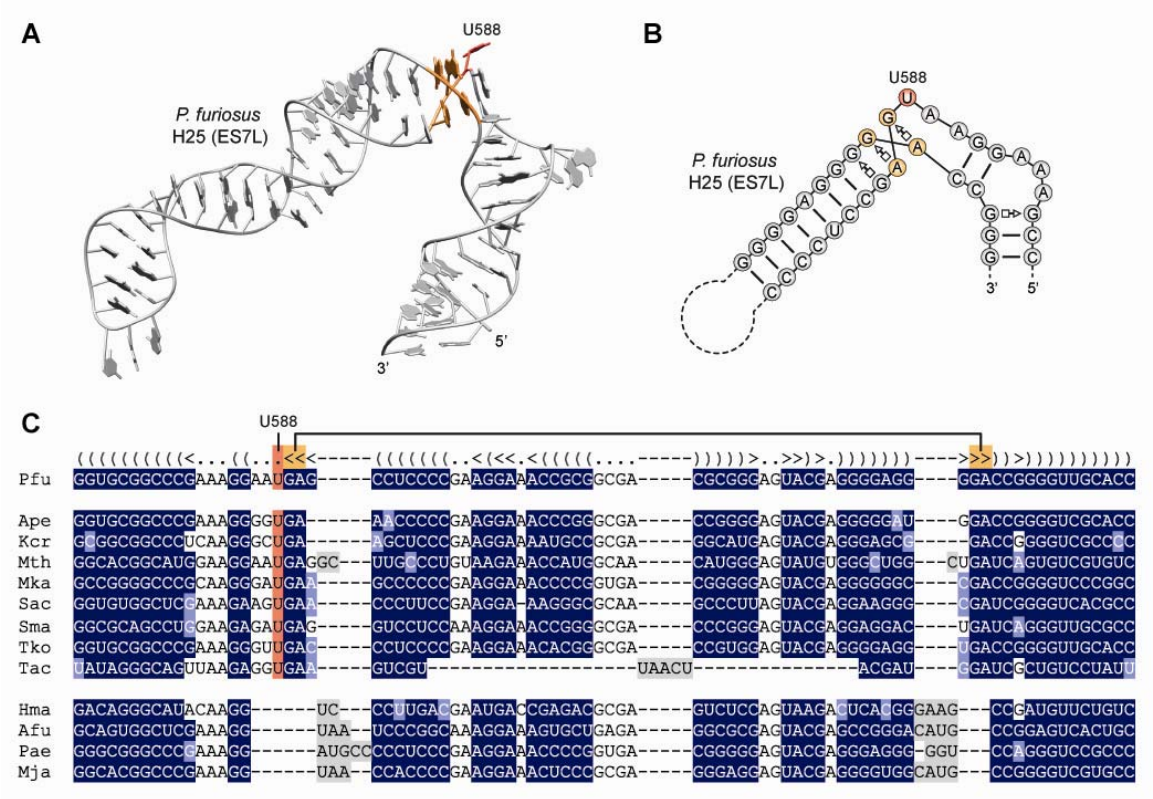
Supplementary Figure 10 Structure and sequence alignments of L8e/L14e/LX with S12e/L27e/L20e. (**A-C**) Superimpositions of the structures of (A) *P. furiosus* L8e with *S. cerevisiae* S12e (7), (B) *P. furiosus* L14e with *S. cerevisiae* L27e (7), and (C) *P. furiosus* LX with *S. cerevisiae* L20e (7). The corresponding sequence alignments shown below were generated using Clustal W (34). Conserved residues are highlighted.



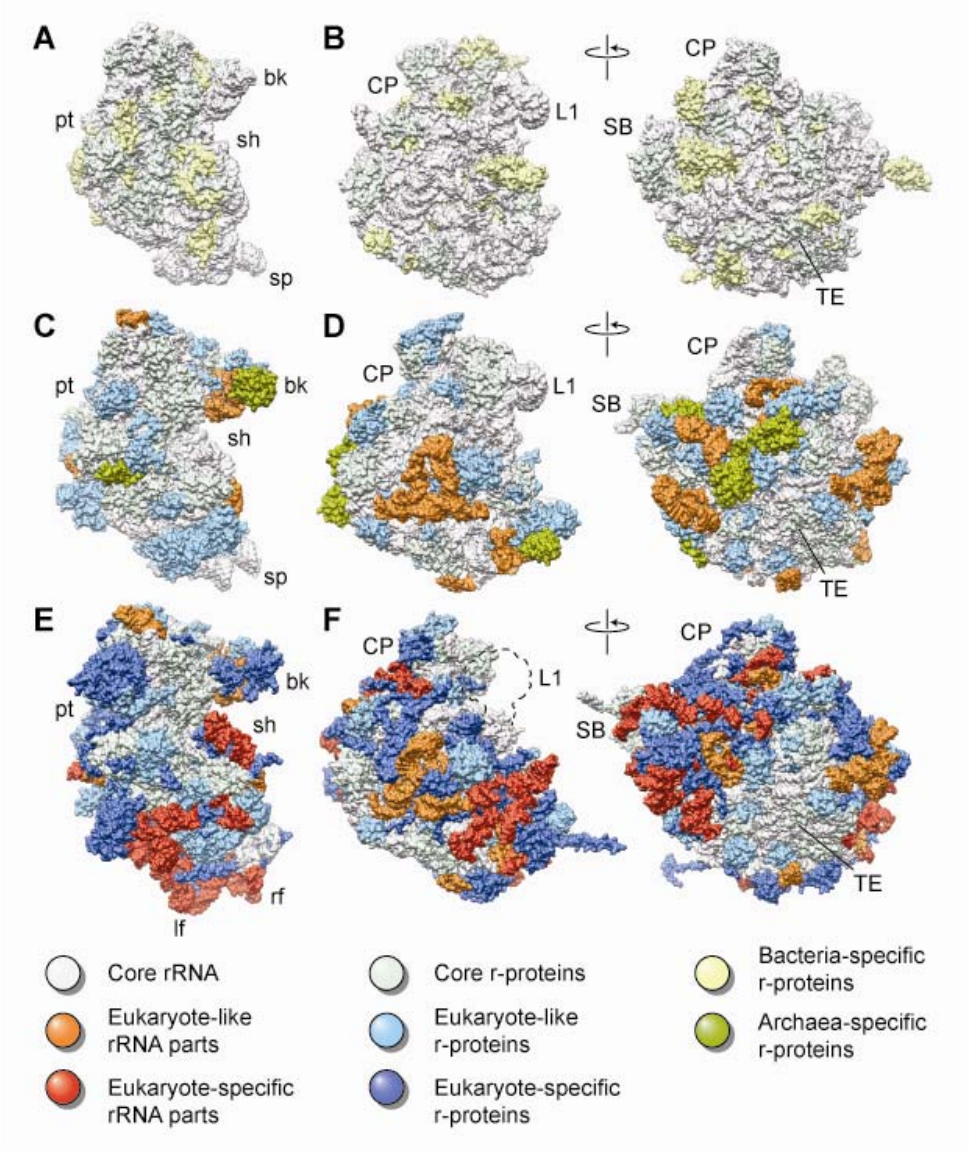
Supplementary Figure 11 Phylogenic distribution of the KT-33 in archaea and eukaryotes. (**A**) Structure of KT-33 with bulged guanine nucleotide (red) from the *S. cerevisiae* ribosome (7). (**B**) Secondary structure diagram of KT-33 from (A). (**C**) Structural alignment of KT-33 showing conservation of guanine in eukaryotes and uridine in archaea (equivalent to position G1228 in *S. cerevisiae* shaded red). The two tandem sheared A/G base pairs characteristic for the KT motif (indicated in orange) are conserved in Archaea and Eukarya. The alignment was generated in S2S (27) using the yeast (Sce) KT-33 structure as template. Colors highlight structural conservation according to the Leontis-Westhof classification (53) implemented in S2S (27), with dark colors indicating isosteric secondary (blue) or tertiary (brown) interactions. Light colors indicate interactions that are geometrically possible, but not necessarily isosteric with the template base-pairs. Nucleotides highlighted in two colors are simultaneously involved in secondary and tertiary interactions. Abbreviations: *Apis mellifera* (Ame), *Caenorhabditis elegans* (Cel), *Danio rerio* (Dre), *Gallus gallus* (Gga), *Homo sapiens* (Has), *Oryza sativa* (Osa), *Saccharomyces cerevisiae* (Sce), *Tetrahymena thermophila* (Tth), *Trypanosoma brucei* (Tbr), *Xenopus laevis* (Xle), *Aeropyrum pernix* (Ape), *Archaeoglobus fulgidus* (Afu), *Haloarcula marismortui* (Hma), *Korarchaeum cryptofilum* (Kcr), *Methanobacterium thermoautotrophicus* (Mth), *Methanococcus jannaschii* (Mja), *Methanopyrus kandleri* (Mka), *Pyrobaculum aerophilum* (Pae), *Pyrococcus furiosus* (Pfu), *Staphylothermus marinus* (Sma), *Sulfolobus acidocaldarius* (Sac), *Thermococcus kodakaraensis* (Tko), *Thermoplasma acidophilum* (Tac).



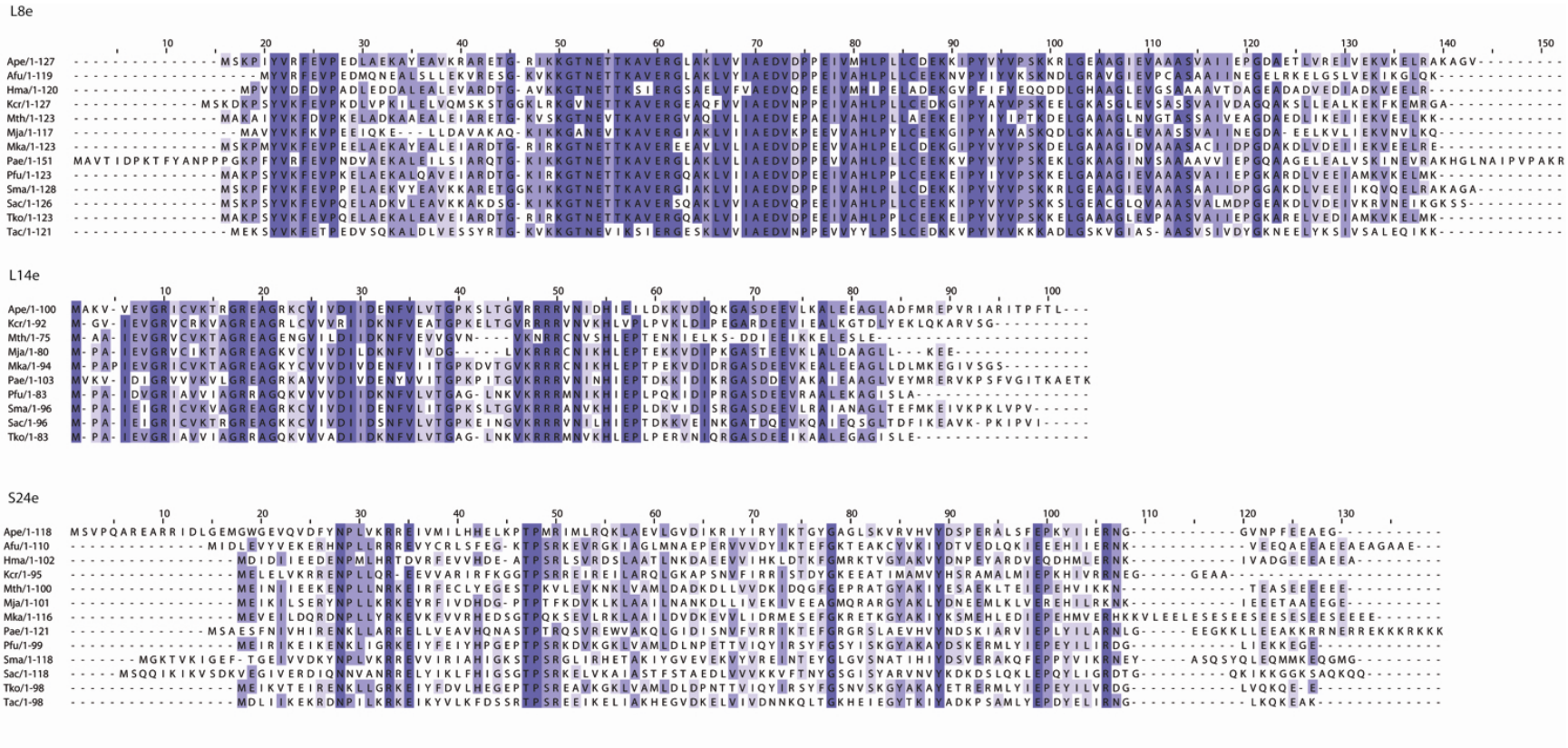
Supplementary Figure 12 Distribution of S24e(L) and L14e(2) binding sites in archaeal ribosomes. (**A-L**) Fit of molecular models of (**A-F**) S24e (blue) and (**G-H**) L14e (green) into the cryo-EM density (mesh) of the (**A,G**) *M. thermautotrophicus* 50S (EMD-2012) (15), (**B,H**) *S. acidocaldarius* 50S (EMD-1797) (21), (**C,I**) *S. marinus* 50S, 50S subunit of the (**D,J**) *T. kodakaraensis* and (**E,K**) *M. igneus* 70S ribosomes, and the (**F,L**) *P. aerophilum* 50S (EMD-1796) (21). Ribosomal RNA is shown in tan with additional r-proteins coloured light blue.



Supplementary Figure 13 Phylogenic distribution of the KT-25 in archaea. (**A**) Structure of KT-25 with bulged uridine nucleotide (red) from the *P. furiosus* ribosome. (**B**) Secondary structure diagram of KT-25 from (A). (**C**) Structural alignment of KT-25 showing conservation of uridine (shaded red, equivalent to position U588 in *P. furiosus*) and KT motif characteristic tandem sheared A/G base pairs (orange). The two KT features can be found in many archaea, but are absent in *Haloarcula marismortui* (Hma), *Archaeoglobus fulgidus* (Afu), *Pyrobaculum aerophilum* (Pae) and *Methanococcus jannaschii* (Mja). The alignment was generated in S2S (27) using the *P. furiosus* (Pfu) KT-25 structure as template. Colors highlight structural conservation according to the Leontis-Westhof classification (53) implemented in S2S (27), with dark colors indicating isosteric secondary (blue) interactions. Non-alignable sequences are shaded grey. Abbreviations used as in Supplementary Figure 11.



Supplementary Figure 14 Intermediate complexity of archaeal ribosomes compared to bacteria and eukaryotes. (**A-F**) Comparison of the structures of the (**A-B**) bacterial (29), (**C-D**) archaeal and (**E,F**) eukaryotic ribosomes (7). Core rRNA (white) and r-proteins (pale green) are conserved in bacteria, archaea and eukaryotes. The bacterial (**A**) small and (**B**) large subunit are shown with bacterial-specific r-proteins coloured yellow, whereas for the archaeal (**C**) small and (**D**) large subunit, the archaeal-specific r-proteins are coloured green, the eukaryotic-like r-proteins and rRNA parts are coloured light blue and orange, respectively. In the eukaryotic 80S ribosome, the eukaryotic-specific r-proteins and rRNA parts are coloured blue and red, respectively.



Supplementary Figure 15 Sequence alignments of archaeal r-proteins L8e, L14e and S24e. Increasing conservation is shown with darker shading of the amino acids (<40%, white, >40%, >60% to >80%). Abbreviations used as in Supplementary Figure 11.

Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution

Jean-Paul Armache[a,1], Alexander Jarasch[a,1], Andreas M. Anger[a,1], Elizabeth Villa[b], Thomas Becker[a], Shashi Bhushan[a], Fabrice Jossinet[c], Michael Habeck[d,e], Gülcin Dindar[a], Sibylle Franckenberg[a], Viter Marquez[a], Thorsten Mielke[f,g], Michael Thomm[h], Otto Berninghausen[a], Birgitta Beatrix[a], Johannes Söding[a], Eric Westhof[c], Daniel N. Wilson[a,2], and Roland Beckmann[a,2]

[a]Gene Center and Center for Integrated Protein Science Munich, Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany; [b]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany; [c]University de Strasbourg, Institut de Biologie Moleculaire et Cellulaire du Centre National de la Recherche Scientifique, 15 Rue René Descartes, 67084 Strasbourg, France; [d]Department of Empirical Inference, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany; [e]Department of Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany; [f]UltraStrukturNetzwerk, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany; [g]Institut für Medizinische Physik und Biophysik, Charité, Ziegelstrasse 5-8, 10117 Berlin, Germany; and [h]Universität Regensburg, Lehrstuhl für Mikrobiologie, Universitätstrasse 31, 93053 Regensburg, Germany

Edited* by Gunter Blobel, The Rockefeller University, New York, NY, and approved September 8, 2010 (received for review July 9, 2010)

Protein biosynthesis, the translation of the genetic code into polypeptides, occurs on ribonucleoprotein particles called ribosomes. Although X-ray structures of bacterial ribosomes are available, high-resolution structures of eukaryotic 80S ribosomes are lacking. Using cryoelectron microscopy and single-particle reconstruction, we have determined the structure of a translating plant (*Triticum aestivum*) 80S ribosome at 5.5-Å resolution. This map, together with a 6.1-Å map of a *Saccharomyces cerevisiae* 80S ribosome, has enabled us to model ∼98% of the rRNA. Accurate assignment of the rRNA expansion segments (ES) and variable regions has revealed unique ES–ES and r-protein–ES interactions, providing insight into the structure and evolution of the eukaryotic ribosome.

modeling | molecular dynamics | flexible fitting

In all living cells, the translation of mRNA into polypeptide occurs on ribosomes. Ribosomes provide a platform upon which aminoacyl-tRNAs interact with the mRNA as well as position the aminoacyl-tRNAs for peptide-bond formation (1). Ribosomes are composed of two subunits, a small subunit that monitors the mRNA–tRNA codon-anticodon duplex to ensure fidelity of decoding (2, 3) and a large subunit that contains the active site where peptide-bond formation occurs (4). Both the small and large subunits are composed of RNA and protein: In eubacteria such as *Escherichia coli*, the small subunit contains one 16S rRNA and 21 ribosomal proteins (r proteins), whereas the large subunit contains 5S and 23S rRNAs and 33 r proteins. Crystal structures of the complete bacterial 70S ribosome were initially reported at 5.5 Å (5), with an interpretation based on atomic models of the individual subunit structures (6–8), and are now available at atomic resolution (9). These structures have provided unparalleled insight into the mechanism of different steps of translation (1) as well as inhibition by antibiotics (10).

Compared to the bacterial ribosome, the eukaryotic counterpart is more complicated, containing expansion segments (ES) and variable regions in the rRNA as well as many additional r proteins and r-protein extensions. Plant and fungal 80S ribosomes contain ∼5,500 nucleotides (nts) of rRNA and ∼80 r proteins, whereas bacterial 70S ribosomes comprise ∼4,500 nts and 54 r proteins. The additional elements present in eukaryotic ribosomes may reflect the increased complexity of translation regulation in eukaryotic cells, as evident for assembly, translation initiation, and development, as well as the phenomenon of localized translation (11–15).

Early models for eukaryotic ribosomes were derived from electron micrographs of negative-stain or freeze-dried ribosomal particles (16) and localization of r proteins was attempted using immuno-EM and cross-linking approaches; see, for example, refs. 17–20. The first cryo-EM reconstruction of a eukaryotic 80S ribosome was reported for wheat germ (*Triticum aestivum*) at 38 Å (21). Initial core models for the yeast 80S ribosome were built at 15-Å resolution (22) by docking the rRNA structures of the bacterial small 30S subunit (6) and archaeal large 50S subunit (8), as well as docking of corresponding homology models of the r proteins. Recently, reconstructions at about 9-Å resolution of fungal and dog 80S ribosomes were used to extend the molecular models to include rRNA expansion segments (23, 24). However, due to the modest resolution, the completeness and accuracy of these models are also limited.

Here we have determined a cryo-EM structure of a wheat germ (*T. aestivum*) translating 80S ribosome at 5.5-Å resolution, enabling us to systematically model ∼98% of the rRNA. This effort encompasses the de novo modeling of 1,885 nucleotides comprising structurally variable regions and rRNA expansion segments. The model reveals direct interaction between ES3 and ES6 as predicted previously by Alkemar and Nygård (25), as well as r-protein–ES interactions, such as L6e and L28e with $ES7^L$ and L34e and L38e with $ES27^L$. The accurate modeling of the rRNA has enabled the localization of 74 (92.5%) of the 80 r proteins of the 80S ribosome (see ref. 26).

Results and Discussion

Cryo-EM Reconstructions of *T. aestivum* and Yeast 80S Ribosomes. Cryo-EM and single-particle analysis were used to reconstruct the *T. aestivum* translating 80S ribosome (Fig. 1*A*) at 5.5-Å resolution (Fig. S1). Similarly, we have previously reported a cryo-EM structure of a translating *Saccharomyces cerevisiae* 80S ribosome

Author contributions: R.B. designed research; J.-P.A., A.J., A.M.A., E.V., T.B., S.B., F.J., M.H., G.D., S.F., V.M., T.M., O.B., B.B., J.S., E.W., and D.N.W. performed research; ; E.V., F.J., M.H., M.T., J.S., and E.W. contributed new reagents/analytic tools; J.-P.A., A.J., A.M.A., E.V., T.B., F.J., E.W., D.N.W., and R.B. analyzed data; and J.-P.A., A.J., A.M.A., D.N.W., and R.B. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: Coordinates of the atomic models of yeast and *Triticum aestivum* 80S complex have been deposited in the Protein Data Bank (PDB), www.pdb.org [PBD ID codes 3IZE, 3IZF, 3IZD (yeast rRNA), 3IZB, 3IZC (yeast r-proteins), and 3IZ7, 3IZ9 (*T. aestivum* rRNA), 3IZ6, 3IZ5 (*T. aestivum* r proteins)]. The cryoelectron microscopic map of the *T. aestivum* 80S-RNCs has been deposited in the 3D-Electron Microscopy Data Bank (EMDB, http://www.ebi.ac.uk/pdbe/emdb/) (EBMD ID code EMD-1780).

[1]J.-P.A., A.J., and A.M.A. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: beckmann@lmb.uni-muenchen.de or wilson@lmb.uni-muenchen.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1009999107/-/DCSupplemental.

at 6.1-Å resolution (Fig. 1*B*) (27). For both reconstructions, ribosome-nascent chain complexes in the posttranslocational state were utilized (27, 28), after in silico sorting (see *Experimental Procedures*) to increase conformational homogeneity. The final reconstruction of the *T. aestivum* 80S ribosome was derived from 1,362,920 particles sorted for the presence of peptidyl-tRNA in the P site (Fig. 1*A*). The resulting cryo-EM map displays characteristics similar to X-ray crystallographic maps of ribosomes at similar resolution, namely, the *Haloarcula marismortui* 50S subunit at 5 Å (29) and the *Thermus thermophilus* 30S (30) and 70S structures (5) at 5.5 Å (Fig. 1 *C*–*F*). At this resolution, well-resolved density for double-helical RNA is observed, with defined minor and major grooves as well as distinctive bumps indicative of phosphate groups located along the backbone ridges (Fig. 1 *C* and *D*). In many regions, single-stranded rRNA sections are traceable and assignment of bulged nucleotides is possible, as reported previously for the 5.8-Å cryo-EM map of TnaC-stalled bacterial 70S ribosome (31). For r proteins, α-helices are observed as rod-like densities (Fig. 1*E*) and β-sheets are represented by smooth surfaces (Fig. 1*F*). The α-helix pitch and strand separation for β-sheets are indiscernible, as expected at this resolution.

Near-Complete Models for the *T. aestivum* and Yeast 80S Ribosomes. The majority of the conserved core of the *T. aestivum* and yeast 80S ribosomes was modeled based on homology of the eukaryotic rRNA with the available bacterial and archaeal ribosome structures using Assemble (32). On this basis, it was possible to generate a template-based model for the *T. aestivum* 80S with a total of 3,466 (1,051/40S and 2,415/60S) nts of the 5485 rRNA, incorporating isosteric base substitutions (33) (Fig. 2). Nucleotides that were not available in the sequences for *T. aestivum* (120 nts, 2.2%) were substituted with those from the closely related *Oryza sativa*. One-hundred sixteen (67 and 49) nts, mostly single-stranded linker regions, could not be modeled (orange in Fig. 2 *A* and *B*; enlargement in Figs. S2–S7) due to unreliable secondary structure predictions and/or ambiguity in the electron density. The remaining 1,903 nts comprising structurally variable regions and rRNA expansion segments were modeled de novo (green in Fig. 2 *A*–*D*) using Assemble (32), RNAfold (34), and RNAshapes (35). Similarly, models for 44 of the 80 r proteins of the *T. aestivum* 80S ribosome (gray in Fig. 2 *E* and *F*) were built using the templates present in the bacterial and archaeal ribosome structures (29, 30), as well as 44 of 79 r proteins of the yeast 80S ribosome (see ref. 26). Models were fitted into the density using Molecular Dynamics Flexible Fitting (MDFF) (36). The *T. aestivum* and yeast 80S models contain all five expansion segments (ES3^S, ES6^S, ES7^S, ES9^S, and ES12^S following the ES nomenclature of ref. 37) and five variable regions (h6, h16, h17, h33, and h41) of the small subunit, as well as the 16 expansion segments (ES3^L, ES4^L, ES5^L, ES7^L, ES9^L, ES10^L, ES12^L, ES15^L, ES19^L, ES20^L, ES24^L, ES26^L, ES27^L, ES31^L, ES39^L, and ES41^L) and two variable regions (H16–18 and H38) of the large subunit (Fig. 3). On the small subunit, the majority of the additional rRNA is clustered at the spur or foot region, except for ES9^S which is positioned at the head (Fig. 3 *C* and *E*). On the large subunit, most ES are located on the back and sides of the particle, leaving the subunit interface and exit tunnel regions essentially unaffected (Fig. 3 *D* and *F*).

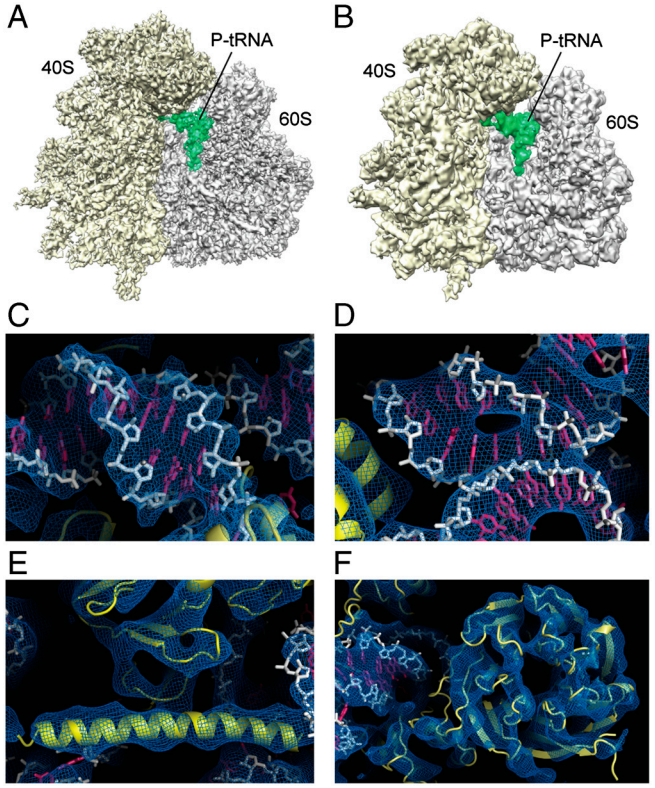


Fig. 1. Cryo-EM reconstruction of eukaryotic 80S ribosomes. (*A*) *T. aestivum* and (*B*) *S. cerevisiae* 80S ribosomes, with small (40S) and large (60S) subunits colored yellow and gray, respectively and the P-tRNA, green. (*C*–*F*) Selected views of the *T. aestivum* 80S density map (blue mesh) and corresponding molecular model, with r protein in yellow and rRNA in white (backbone) and red (bases).

Comparison of Expansion Segments Between Yeast and Wheat Germ. Interestingly, the density maps of the *T. aestivum* and yeast 80S reconstructions support a direct interaction of the loop of ES6^{S}d with the internal loop between ES3^{S}b and ES3^{S}c (Fig. 4*A*), as predicted previously by Alkemar and Nygård (25). Although this interaction was not modeled in the fungal or canine 80S ribosomes (23, 24), covariation analysis suggests that the ES3^S–ES6^S base-pairing interaction is also conserved in mammalian 80S ribosomes (25).

The largest ES in the *T. aestivum* and yeast ribosomes is ES7^L, which is located at the back of the 60S subunit (Fig. 4*B*). Overall, ES7^L is similar between *T. aestivum* and yeast, however, at least two clear differences are evident: Firstly, density for ES7^{L}a in yeast is only seen at low thresholds (Fig. S8), suggesting it is more flexible than in *T. aestivum* (Fig. 4*B*). The reason for this flexibility appears to be that ES7^{L}a in *T. aestivum* is stabilized through an interaction with r-protein L28e (Fig. 4*B*), which is not present in the *S. cerevisiae* genome (38). Secondly, *T. aestivum* contains a three-way junction formed by ES7^{L}c–e, whereas this architecture is not present in yeast due to the absence of ES7^{L}d,e (Fig. 4*B* and Figs. S2–S7). Surprisingly, the N terminus of *T. aestivum* r-protein L6e, which is shorter in yeast, appears to insert through the three-way junction of ES7^L (Fig. 4*B*), an RNA-protein interaction that has to our knowledge not been reported previously.

ES27^L is unique for its highly dynamic behavior, being found in two distinct positions in yeast 80S ribosomes (39); one oriented toward the L1 stalk, termed ES27$^L_{in}$ and one away from the L1 stalk but toward the tunnel exit, termed ES27$^L_{out}$ (Fig. 4*C*). Modeling of both conformations reveals that interchange between the ES27$^L_{in}$ (gold) and ES27$^L_{out}$ (blue) positions involves a rotation of ~110° of ES27^{L}a–c relative to H63 (Fig. 4*C*). Weak density for ES27^L in the reconstruction of the *T. aestivum* ribosome suggests that ES27^L exhibits a continuum of different conformational states. Nevertheless, at low thresholds, one preferential state is observed, intermediate in position (ES27$^L_{int}$) to the yeast ES27$^L_{in}$ and ES27$^L_{out}$ positions (Fig. 4*C* and Fig. S8). All three positions appear to be stabilized through interaction with newly identified eukaryotic-specific r proteins: The yeast ES27$^L_{out}$ and the *T. aestivum* ES27$^L_{int}$ conformations directly contact r-protein L38e (Fig. 4*C*), whereas r-protein L34e stabilizes the yeast ES27$^L_{in}$ position. In *Tetrahymena*, deletion of

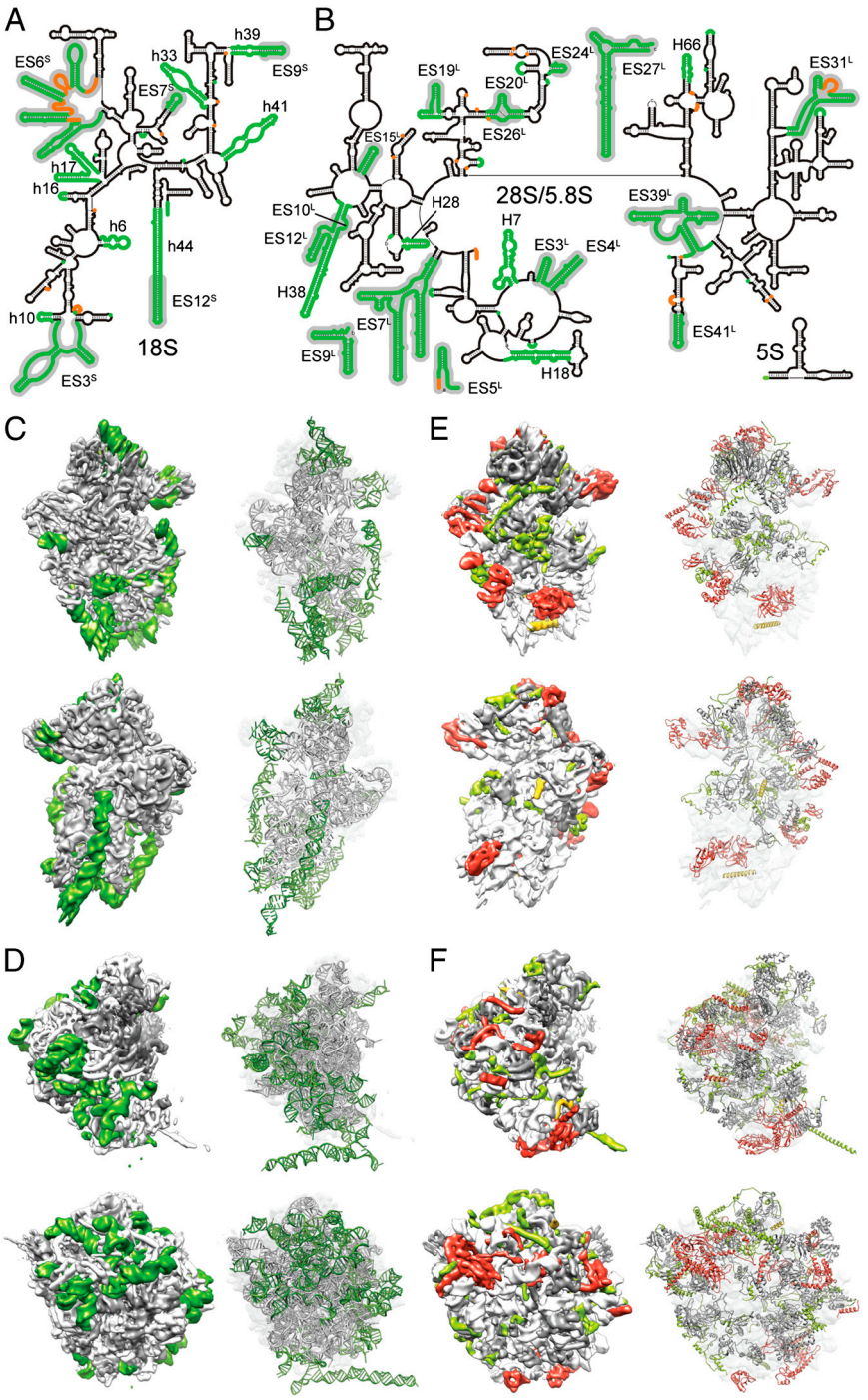


Fig. 2. An atomic model for the *T. aestivum* 80S ribosome. (*A* and *B*) Secondary structures for the (*A*) small (18S) and (*B*) large subunit (5S, 5.8S, and 28S) ribosomal RNAs, with the newly modeled regions colored green. Expansion segments and variable regions are indicated in gray and unmodeled regions are orange. (*C* and *D*) Newly modeled regions of rRNA (green) are highlighted on the (*C*) small and (*D*) large subunit density map (*Left*) and as molecular models (*Right*). (*E* and *F*) Newly modeled proteins are highlighted on the (*E*) small and (*D*) large subunit density map (*Left*) and as molecular models (*Right*). Newly identified proteins are colored red, whereas de novo modeled extensions are colored light green, and modeled but unassigned proteins are yellow.

ES27^{L} is lethal (40), suggesting a functionally important role for this RNA insertion. Despite the high variability in length of ES27^{L}, ranging from ∼150 nucleotides in *T. aestivum* and yeast to ∼700 nucleotides in mammals (41), the ES27^{L} deletion can be complemented with a corresponding ES27^{L} from other species (40). ES27^{L} has been suggested to play a role in coordinating the access of nonribosomal factors at the tunnel exit (39).

Evaluation of RNA Models for the Eukaryotic Ribosome. A reconstruction at 8.7 Å of a canine ribosome was used for a model including models of ∼50% of the ES by fitting of A-form helices into the ES density (23). Recently, a more comprehensive model of the yeast *S. cerevisiae* ribosome was built on the basis of an 8.9-Å cryo-EM reconstruction of a 80S ribosome from a related thermophilic fungus, *Thermomyces lanuginosus* (24), which, however, shares only ∼85% sequence identity with *S. cerevisiae* rRNA. With the exception of ES10^{L}, ES27^{L}, and the tip of ES15^{L}, molecular models were built for all the remaining expansion segments and variable regions (24). Yet, a number of significant differences between the yeast model presented by Taylor et al. (24) and the

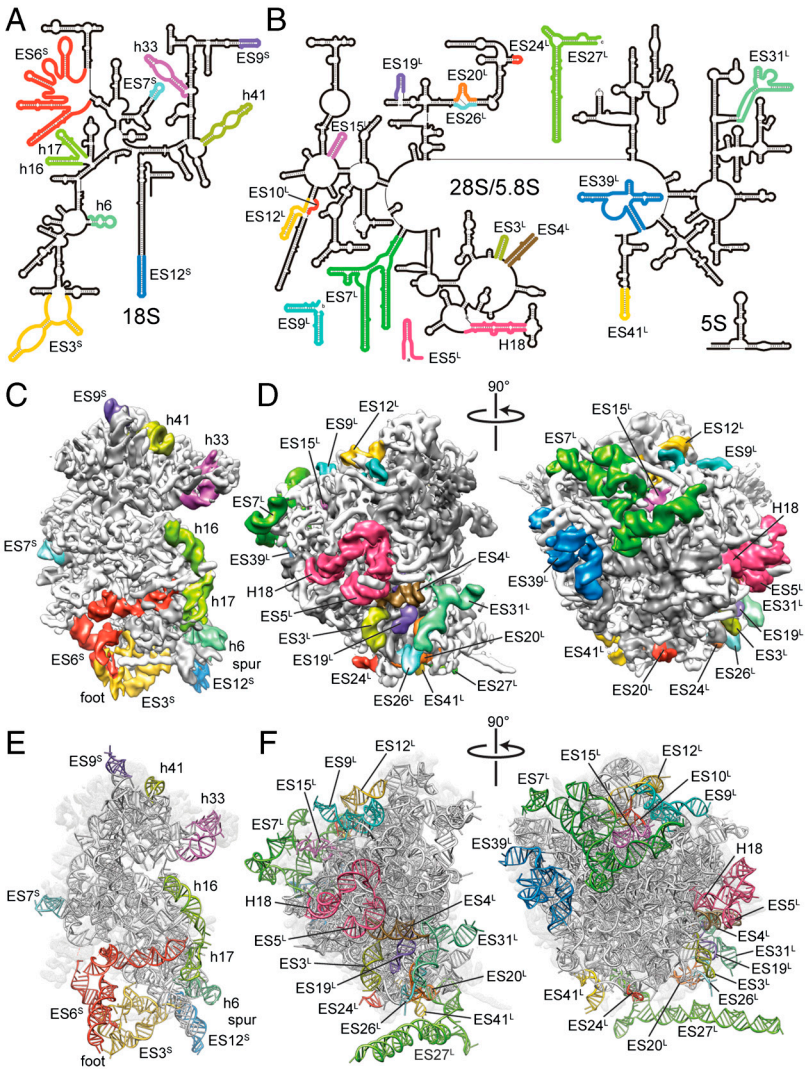


Fig. 3. Ribosomal RNA expansion segments and variable regions. (*A* and *B*) Secondary structures for the *T. aestivum* (*A*) small (18S) and (*B*) large subunit (5S, 5.8S, and 28S) ribosomal RNAs, with the ES and variable regions (VR) colored distinctly. (*C* and *D*) Cryo-EM maps of the (*C*) small and (*D*) large subunits with assigned ES and VR colored as in *A* and *B*. (*E* and *F*) Molecular models of the ES and VR of rRNA colored as in *C* and *D*.

yeast model presented here are evident (Fig. S9). Taken together, using the correct sequences for modeling into corresponding maps of improved resolution allowed for a significant improvement in completeness and accuracy of both the RNA and protein models.

Evolution of RNA Expansion Segments. A comparison of genomic sequences from diverse organisms, ranging from bacteria to mammals, indicates additional mass with increasing organism complexity (Fig. 5). However, the composition of mammalian ribosomes, e.g., from human, is surprisingly similar to those of other eukaryotes, such as yeast and plants described here. Human ribosomes have the same 80 r proteins that are found in *T. aestivum* ribosomes and, in terms of rRNA, differ significantly only in the length of four ES on the large subunit ($ES7^L$, $ES15^L$, $ES27^L$, and $ES39^L$). These are longer in human (~850, ~180, ~700, and ~220 nts) than in *T. aestivum*/yeast (~200, ~20, ~150, and ~120 nts, respectively), and cryo-EM reconstructions of mammalian ribosomes (23, 42–44) show that the longer ES in mammalian ribosomes are generally highly mobile elements for which little to no density is visible (Fig. 5). Evolution has thus favored the development of two apparently distinct layers of mass gain for the ribosome: A first layer of tightly intertwined additional proteins and rRNA expansions rigidly positioned on the subunit surfaces (with the only exception of the mobile $ES27^L$), which was followed by a second layer comprising a few drastically extended highly mobile rRNA elements with hitherto unknown function.

Experimental Procedures

Sample Preparation and Cryo-EM. The cryo-EM map used for modeling of the yeast 80S ribosome was previously deposited in the Electron Microscopy Data Bank (EMDB ID 1669; ref. 27). *T. aestivum* ribosome nascent chain complexes (RNCs) were generated using a homemade wheat germ in vitro translation system (based on ref. 45) and were purified according to ref. 46. As described previously (47), *T. aestivum* RNC samples were applied to carbon-coated holey grids, and micrographs were recorded under low-dose conditions on a Polara field emission gun electron microscope at 300 kV (FEI) in a defocus range of 1.0–4.5 μm. Micrographs were scanned on a Heidelberg Primescan D8200 drum scanner, resulting in a pixel size of 1.24 Å on the object scale. The data were analyzed by determination of the contrast transfer function using CTFFIND (48). The data were further processed with the SPIDER software package (49). After automated particle picking followed by visual inspection, 2,108,230 particles of the *T. aestivum* RNC dataset were sorted in a supervised manner (50) into programmed (with

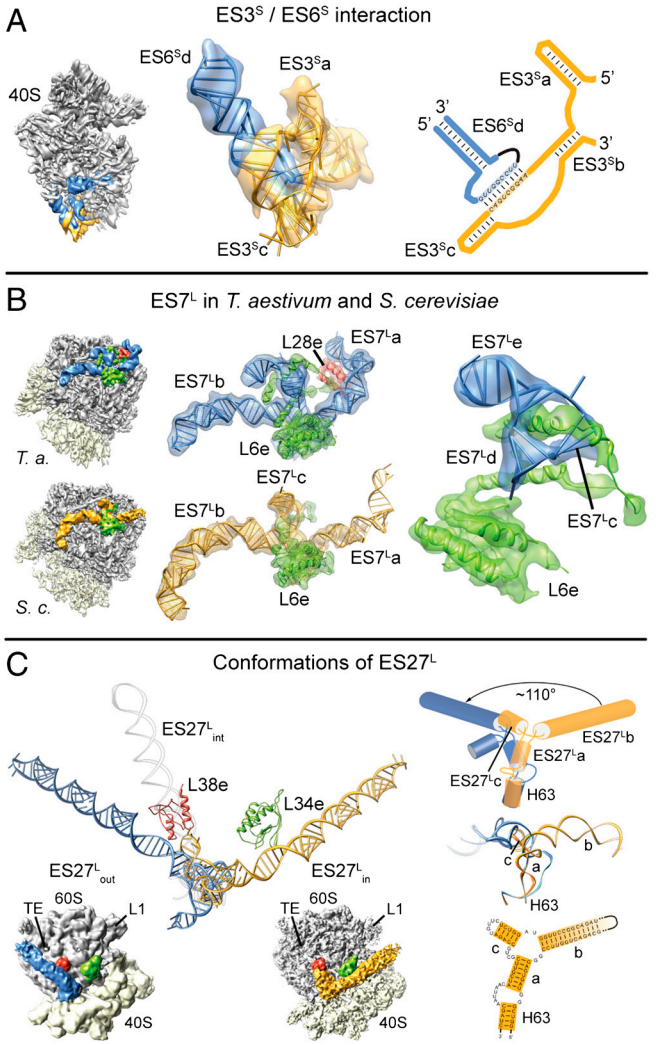


Fig. 4. Molecular models for expansion segments ES3^S/ES6^S, ES7^L, and ES27^L. (*A*) Isolated density for ES6^{S}d (blue) and ES3^{S}a,c (gold) on the 40S subunit (*Left*) and transparent with a molecular model (*Center*). rRNA secondary structure prediction highlighting interaction between the loop of ES6^{S}d and the bulge in ES3^{S}c (*Right*), as proposed by ref. 59. (*B*) Isolated density for ES7^L from *T. aestivum* (*T. a.*, blue) and *S. cerevisiae* (*S. c.*, gold) on the 80S ribosome (*Left*) and transparent with a molecular model (*Center*). Ribosomal proteins L28e (red) stabilizes ES7^{L}a in the *T. aestivum* 80S ribosome, whereas the extension of r-protein L6e appears to pass through the three-way junction formed by helices ES7^{L}c–e (*Right*). Molecular models for the ES27$^L_{in}$ (gold) and ES27$^L_{out}$ (blue) positions (*Left*), as observed in *S. cerevisiae* 80S ribosomes (*Thumbnail Insets*) (39) and an intermediate position (ES27$^L_{int}$, gray) observed in the *T. aestivum* 80S ribosome. In yeast, r-protein L34e (green) and L38e (red) interact with the ES27$^L_{in}$ and ES27$^L_{out}$ positions, respectively. The tunnel exit (TE) and L1 stalk (L1) are indicated for reference. (*C*) Schematic (*Top Right*) and molecular model (*Middle Right*) indicating that the interchange between the ES27$^L_{in}$ (gold) and ES27$^L_{out}$ (blue) positions involves a rotation of ~110° of ES27^{L}a–c relative to H63. Secondary structure for the junctions of *S. cerevisiae* ES27^{L}a–c and H63.

P-tRNA) and unprogrammed/empty (without P-tRNA) ribosome subdatasets, using reconstructions of programmed and unprogrammed ribosomes as initial references. Removal of unprogrammed ribosome particles resulted in 1,362,920 particles that were used for reconstruction of the wheat germ 80S ribosome. The final contrast transfer function corrected reconstruction has a resolution of 5.5 Å, based on the Fourier Shell Correlation with a cutoff value of 0.5 (Fig. S1). Densities for the 40S subunit, the 60S subunit, and the P-site tRNA were isolated using binary masks.

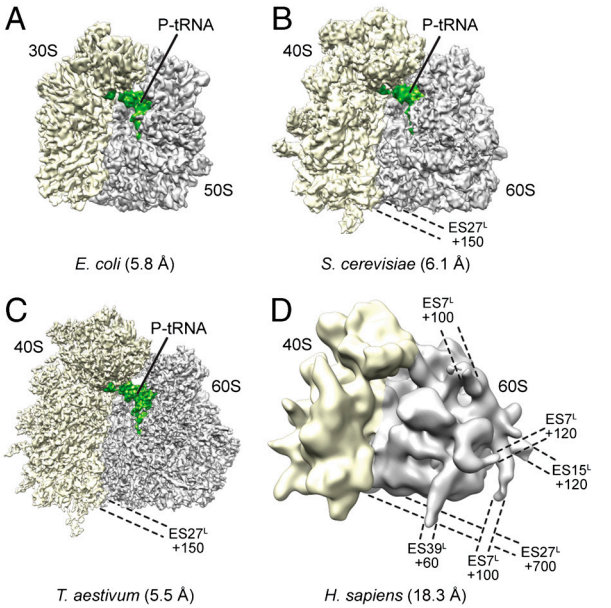


Fig. 5. Cryo-EM reconstructions of ribosomes from (*A*) the eubacterium *Escherichia coli* (31), (*B*) the yeast *S. cerevisiae* (27), (*C*) wheat germ *T. aestivum* (this work), and (*D*) *Homo sapiens* (44). The small and large subunits are shown in yellow and gray, respectively and the P-tRNA (green) is indicated for reference. The dashed lines and numbers indicate the number of nucleotides of the rRNA expansion segments that are not visualized.

Ribosomal RNA Sequences. The rRNA sequences of the *S. cerevisiae* 5S, 5.8S, 18S, and 25S were taken from GenBank Accession number (Acc.) U53879. The rRNA sequence for the *T. aestivum* 5S (Acc. X06094), 5.8S (Acc. FM998894), 18S (Acc. AY049040), and 28S (Acc. AY049041) rRNAs were available, with the exception of five and four nucleotides at the 5′ end of the 18S and 28S, respectively, 46 nts from the 3′ end of the 28S, and 65 nts (487–551) in the 28S corresponding to ES7, which were filled with the corresponding sequences of *O. sativa* (Acc. M11585). Sequence alignments between the available *T. aestivum* and *O. sativa* rRNAs show a 98% sequence identity, indicating the suitability of using the *O. sativa* sequence for filling the missing 120 (2.2%) nucleotides in the *T. aestivum* model.

Modeling of the Ribosomal RNA Core. The structure-based sequence alignment of both the 18S of the small subunit and the 5S, 5.8S, and 28S rRNA of the large subunit was done using the X-ray structure of the large ribosomal of *H. marismortui* [Protein Data Bank (PDB) 1FFK] (8) and the small ribosomal subunit of *T. thermophilus* (PDB 1J5E) (6). For regions like H5–H7, the stalk base (H42–H43), and the L1 stalk (H76–H78), the X-ray structure of *E. coli* (PDB 3FIK) (51, 52) was used as template. The alignment was constructed semiautomatically using S2S (53). The multiple sequence alignment for the 5S, 5.8S, and 28S was constructed between *H. marismortui*, *T. aestivum*/*O. sativa*, and *S. cerevisiae* and for the 18S between *T. thermophilus*, *T. aestivum*/*O. sativa*, and *S. cerevisiae*, respectively. The resulting core models for *S. cerevisiae* and *T. aestivum* were deduced from the alignments using Assemble (32) and core models consist only of isosteric base substitutions (33, 54).

Modeling of the Ribosomal RNA Expansion Segments. Primary sequences were used as an input for RNA secondary structure prediction tools RNAfold (34) and RNAshapes (35). The core model was used as an anchor point for modeling the ES. According to the secondary structure predictions and the electron density, the ES were constructed semiautomatically using Assemble (32). The applied structural motifs for loops and inner helical

non-Watson-Crick base-pairing motifs were extracted from known structures of the PDB and Structural Classification of RNA database (55).

Refinement and Fitting of the rRNA and r-Proteins into the EM Densities. The de novo modeled RNA parts were initially refined using the internal refinement tool of Assemble. A preliminary rigid body fitting of the models was done without proteins using Chimera (56) with low-pass filtered electron densities. Subsequently, all RNA segments were merged using visual molecular dynamics (VMD) (57), and MDFF (36) was applied to fit the rRNA to the density map while preserving canonical and noncanonical base-pair interactions identified by RNAview. Subsequently, proteins were introduced using VMD, and an extended version of interactive molecular dynamics (58), namely, interactive MDFF, was used to refine the proteins into the density while fixing protein–RNA and protein–protein clashes, followed by an MDFF refinement of the entire 80S model.

Visualization and Figure Preparation. Cryo-EM maps and models were visualized and all figures were generated using VMD (57), Chimera (56), and/or PyMol (http://www.pymol.org).

ACKNOWLEDGMENTS. This research was supported by grants from the Deutsche Forschungsgemeinschaft SFB594 and SFB646 (to R.B.), SFB740 (to T.M.), and WI3285/1-1 (to D.N.W.), by the European Union and Senatsverwaltung für Wissenschaft, Forschung und Kultur Berlin (UltraStructureNetwork, Anwenderzentrum), and a Marie Curie International Incoming Fellowship within the Seventh European Community Framework Programme (E.V.). Computer time for MDFF was provided by the Leibniz-Rechenzentrum.

1. Schmeing TM, Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461:1234–1242.
2. Ogle JM, Ramakrishnan V (2005) Structural insights into translational fidelity. *Annu Rev Biochem* 74:129–177.
3. Zaher HS, Green R (2009) Fidelity at the molecular level: Lessons from protein synthesis. *Cell* 136:746–762.
4. Simonovic M, Steitz TA (2009) A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. *Biochim Biophys Acta* 1789:612–623.
5. Yusupov MM, et al. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* 292:883–896.
6. Wimberly BT, et al. (2000) Structure of the 30S ribosomal subunit. *Nature* 407:327–339.
7. Schluenzen F, et al. (2000) Structure of functionally activated small ribosomal subunit at 3.3 A resolution. *Cell* 102:615–623.
8. Ban N, et al. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289:905–920.
9. Selmer M, et al. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313:1935–1942.
10. Wilson DN (2009) The A-Z of bacterial translation inhibitors. *Crit Rev Biochem Mol Biol* 44:393–433.
11. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: Mechanisms and biological targets. *Cell* 136:731–745.
12. Warner JR, McIntosh KB (2009) How common are extraribosomal functions of ribosomal proteins? *Mol Cell* 34:3–11.
13. Freed EF, Bleichert F, Dutca LM, Baserga SJ (2010) When ribosomes go bad: Diseases of ribosome biogenesis. *Mol Biosyst* 6:481–493.
14. Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11:113–127.
15. Wang DO, Martin KC, Zukin RS (2010) Spatially restricting gene expression by local translation at synapses. *Trends Neurosci* 33:173–182.
16. Lake JA (1985) Evolving ribosome structure: Domains in archaebacteria, eubacteria, eocytes and eukaryotes. *Annu Rev Biochem* 54:507–530.
17. Gross B, Westermann P, Bielka H (1983) Spatial arrangement of proteins within the small subunit of rat liver ribosomes studied by cross-linking. *EMBO J* 2:255–260.
18. Marion MJ, Marion C (1987) Localization of ribosomal proteins on the surface of mammalian 60S ribosomal subunits by means of immobilized enzymes. Correlation with chemical cross-linking data. *Biochem Biophys Res Commun* 149:1077–1083.
19. Lutsch G, et al. (1990) Immunoelectron microscopic studies on the location of ribosomal proteins on the surface of the 40S ribosomal subunit from rat liver. *Eur J Cell Biol* 51:140–150.
20. Pisarev AV, et al. (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* 27:1609–1621.
21. Verschoor A, Srivastava S, Grassucci R, Frank J (1996) Native 3D structure of eukaryotic 80S ribosome: Morphological homology with the *E. coli* 70S ribosome. *J Cell Biol* 133:495–505.
22. Spahn CM, et al. (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit–subunit interactions. *Cell* 107:373–386.
23. Chandramouli P, et al. (2008) Structure of the mammalian 80S ribosome at 8.7 A resolution. *Structure* 16:535–548.
24. Taylor DJ, et al. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure* 17:1591–1604.
25. Alkemar G, Nygård O (2006) Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. *Biochemistry* 45:8067–8078.
26. Armache J-P, et al. (2010) Localization of eukaryote-specific ribosomal proteins: Implications for structure, function and evolution. *Proc Natl Acad Sci USA* 10.1073/pnas.1010005107.
27. Becker T, et al. (2009) Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* 326:1369–1373.
28. Bhushan S, et al. (2010) α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* 17:313–317.
29. Ban N, et al. (1999) Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature* 400:841–847.
30. Clemons WMJ, et al. (1999) Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* 400:833–840.
31. Seidelt B, et al. (2009) Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* 326:1412–1415.
32. Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* 26:2057–2059.
33. Stombaugh J, Zirbel CL, Westhof E, Leontis NB (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* 37:2294–2312.
34. Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31:3429–3431.
35. Steffen P, et al. (2006) RNAshapes: An integrated RNA analysis package based on abstract shapes. *Bioinformatics* 22:500–503.
36. Trabuco LG, et al. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16:673–683.
37. Gerbi SA (1996) *Ribosomal RNA—Structure, Evolution, Processing, and Function in Protein Synthesis*, eds RA Zimmermann and AE Dahlberg (CRC, Boca Raton, FL), pp 71–87.
38. Lecompte O, et al. (2002) Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. *Nucleic Acids Res* 30:5382–5390.
39. Beckmann R, et al. (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* 107:361–372.
40. Sweeney R, Chen LH, Yao MC (1994) An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. *Mol Cell Biol* 14:4203–4215.
41. Cannone JJ, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:1–31.
42. Dube P, et al. (1998) The 80S rat liver ribosome at 25 A resolution by electron cryomicroscopy and angular reconstitution. *Structure* 6:389–399.
43. Morgan DG, et al. (2000) A comparison of the yeast and rabbit 80S ribosome reveals the topology of the nascent chain exit tunnel, inter-subunit bridges and mammalian rRNA expansion segments. *J Mol Biol* 301:301–321.
44. Spahn CM, et al. (2004) Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes; the IRES functions as an RNA-based translation factor. *Cell* 118:465–475.
45. Erickson AH, Blobel G (1983) Cell-free translation of messenger RNA in a wheat germ system. *Method Enzymol* 96:38–50.
46. Halic M, et al. (2004) Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* 427:808–814.
47. Wagenknecht T, et al. (1988) Direct localization of the tRNA-anticodon interaction site on the *Escherichia coli* 30S ribosomal subunit by electron microscopy and computerized image averaging. *J Mol Biol* 203:753–760.
48. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142:334–347.
49. Frank J, et al. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116:190–199.
50. Penczek PA, Frank J, Spahn CM (2006) A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J Struct Biol* 154:184–194.
51. Schuwirth B, et al. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310:827–834.
52. Villa E, et al. (2009) Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci USA* 106:1063–1068.
53. Jossinet F, Westhof E (2005) Sequence to structure (S2S): Display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* 21:3320–3321.
54. Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512.
55. Tamura M, et al. (2004) SCOR: Structural classification of RNA, version 2.0. *Nucleic Acids Res* 32:182D–184.
56. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
57. Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. *J Mol Graphics* 14:33–38.
58. Grayson P, Tajkhorshid E, Schulten K (2003) Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics. *Biophys J* 85:36–48.
59. Alkemar G, Nygård O (2003) A possible tertiary rRNA interaction between expansion segments ES3 and ES6 in eukaryotic 40S ribosomal subunits. *RNA* 9:20–24.

Supporting Information

Armache et al. 10.1073/pnas.1009999107

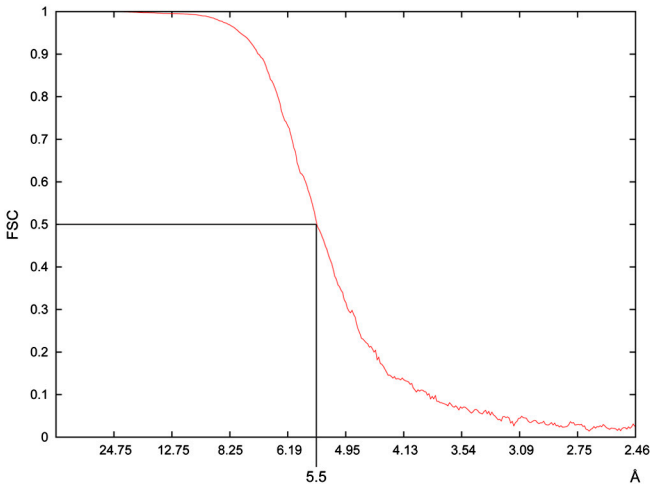


Fig. S1. Fourier shell correlation (FSC) curve for the cryo-EM reconstruction of the *Triticum aestivum* 80S ribosome-nascent chain complex, with resolution of 5.5 Å according to a cut-off of the FSC at 0.5.

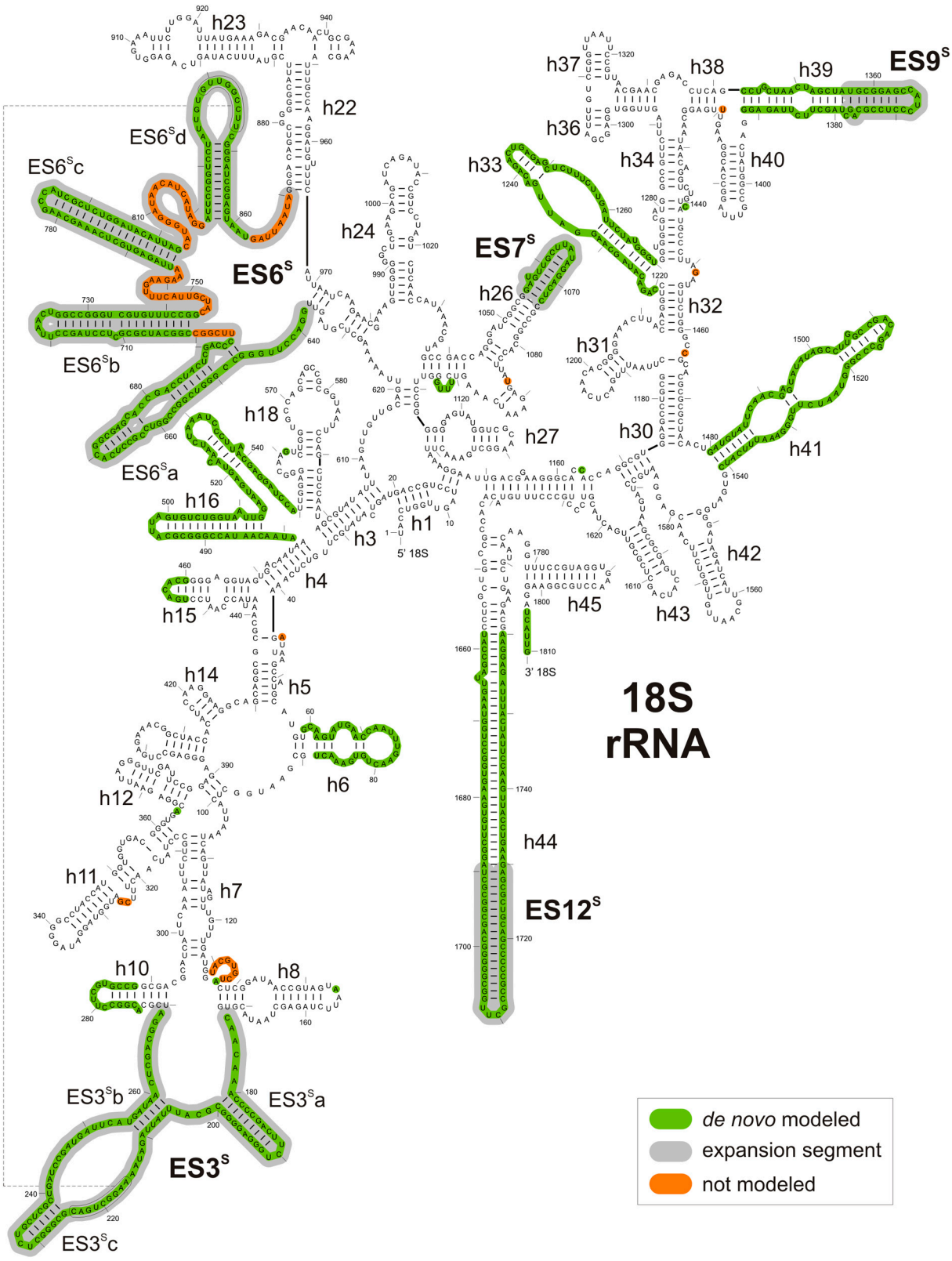


Fig. S2. Secondary structure diagram for the small subunit (18S) rRNA of *Triticum aestivum*, modified from ref. 1. Green regions indicated de novo modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled.

1 Cannone JJ, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

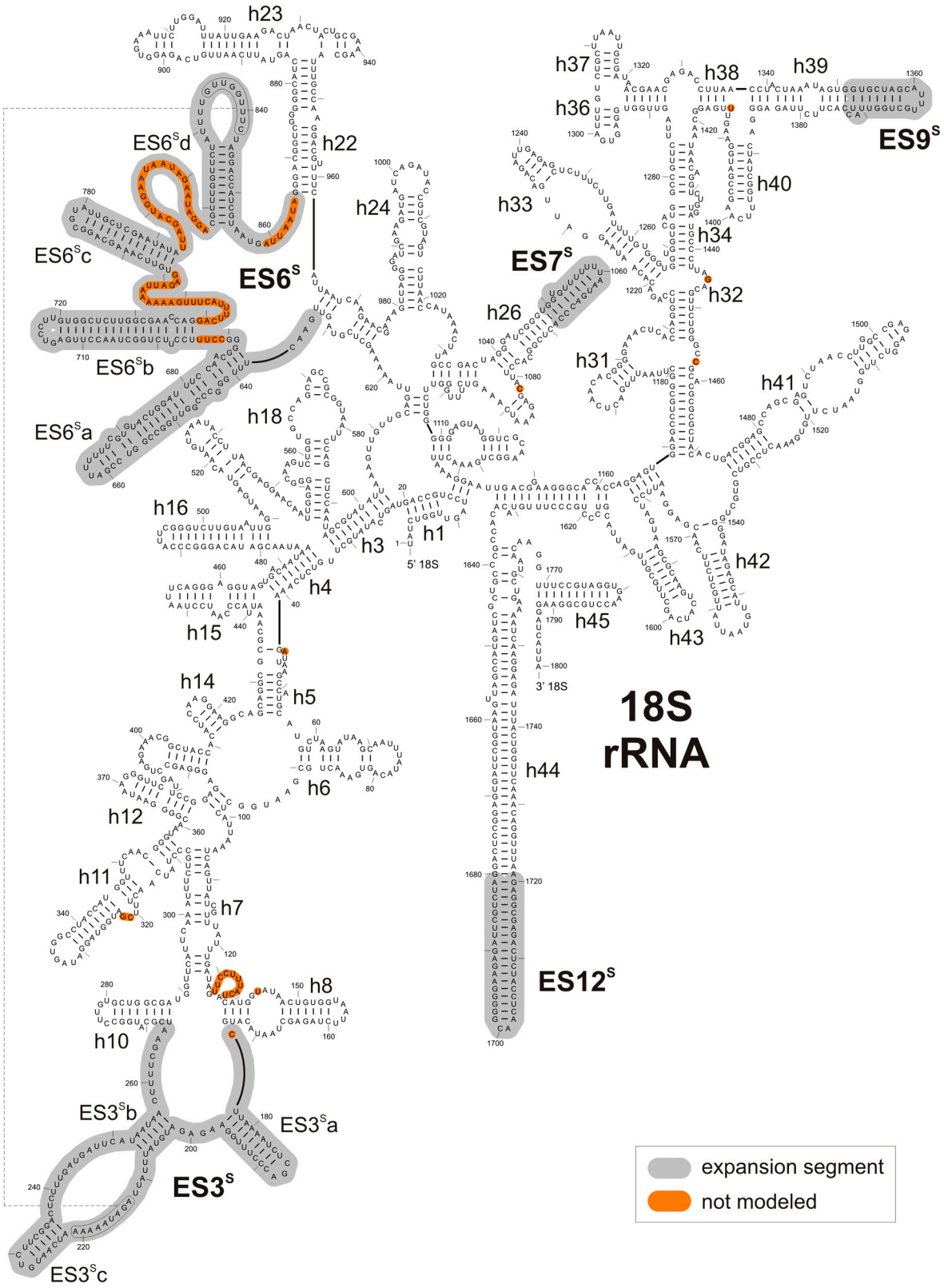


Fig. S3. Secondary structure diagram for the small subunit (18S) rRNA of *Saccharomyces cerevisiae* modified from ref. 1. Gray regions indicate expansion segments, whereas orange nucleotides were not modeled.

1 Cannone JJ, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

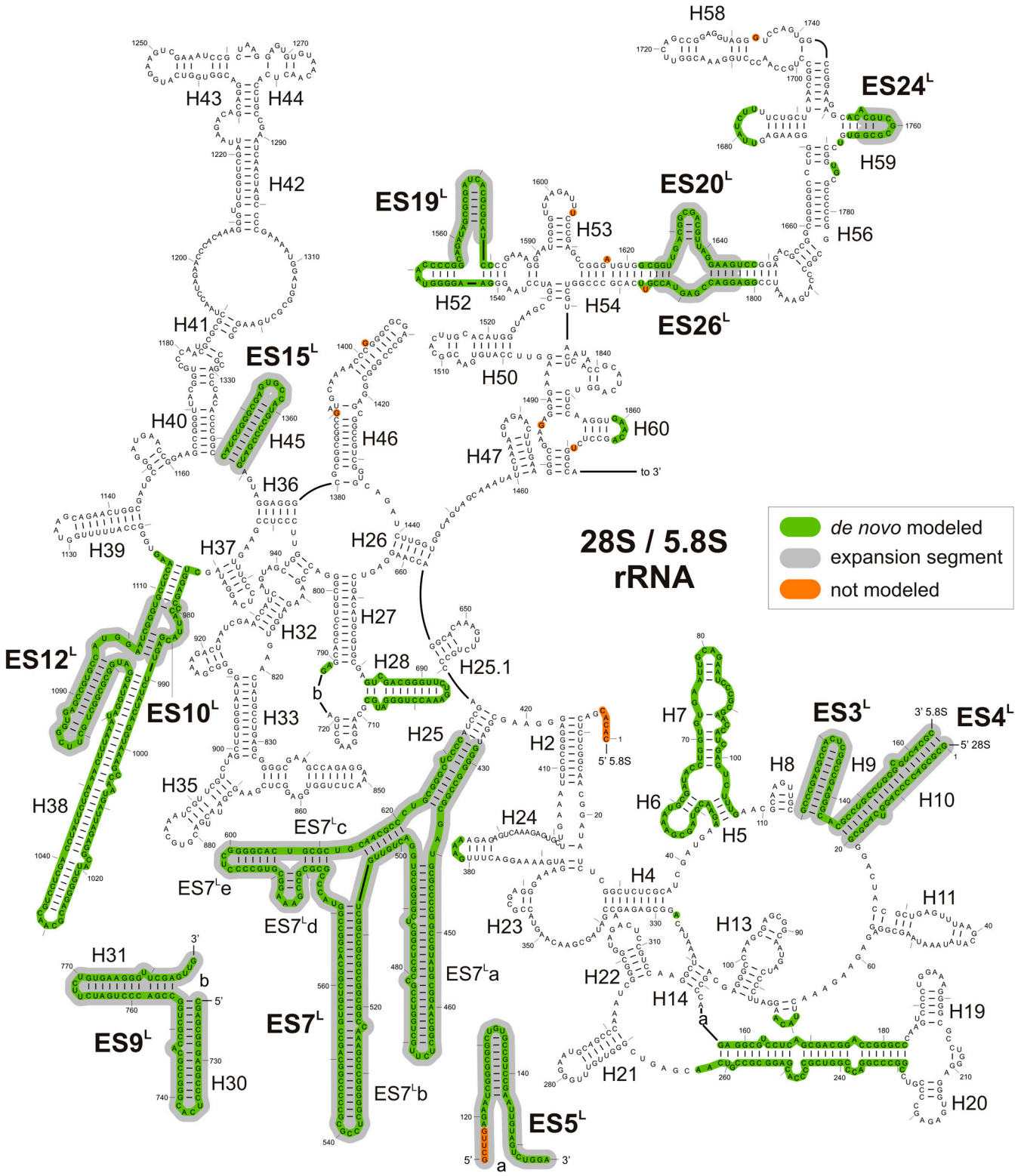


Fig. S4. Secondary structure diagram for the 5′ region of the large subunit rRNAs (5.8S and 28S) of *Triticum aestivum* modified from ref. 1. Green regions indicated de novo modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled.

1 Cannone JJ, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

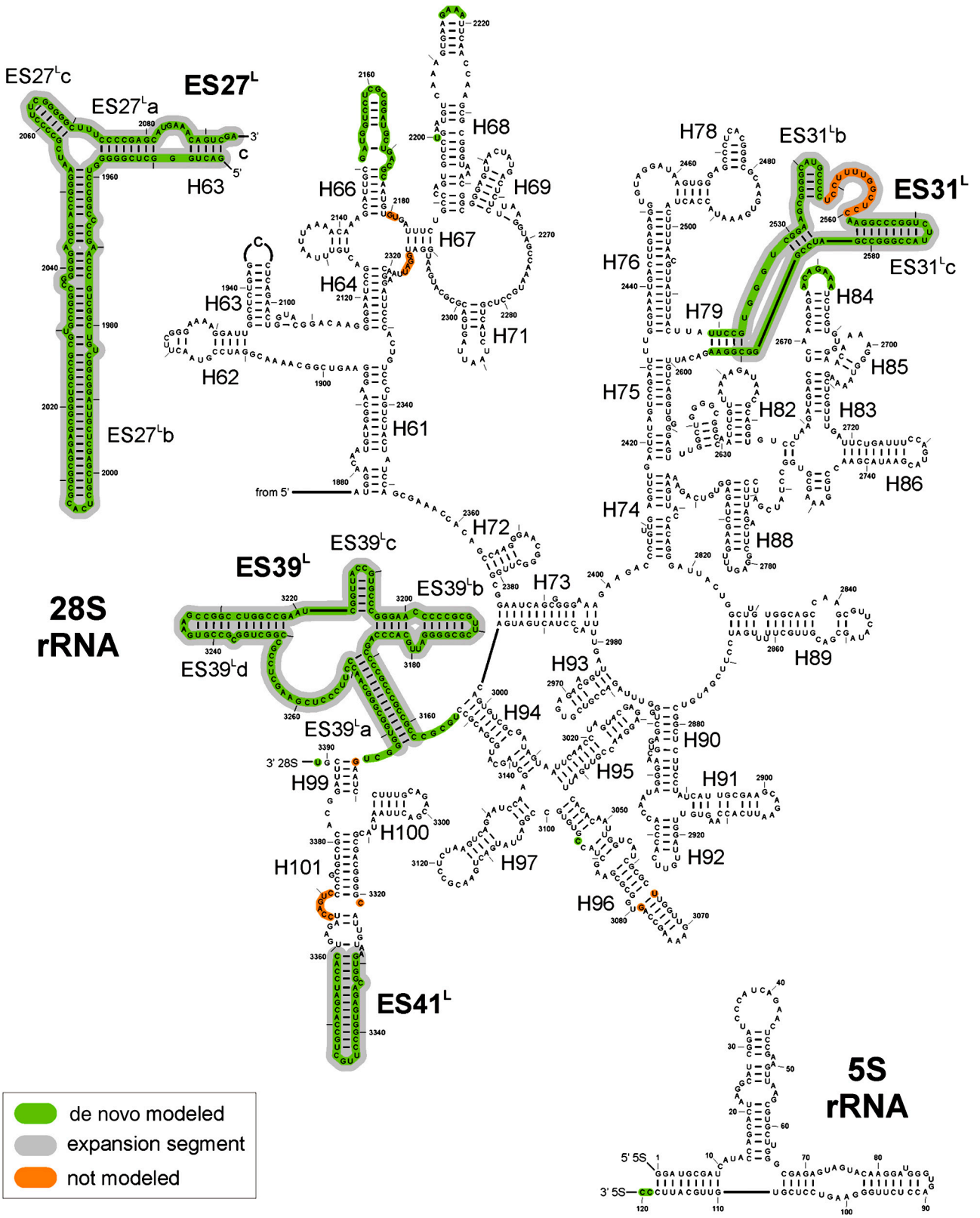


Fig. S5. Secondary structure diagram for the 3′ region of the large subunit rRNAs (5S and 28S) of *Triticum aestivum* modified from ref. 1. Green regions indicated de novo modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled.

1 Cannone JJ, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

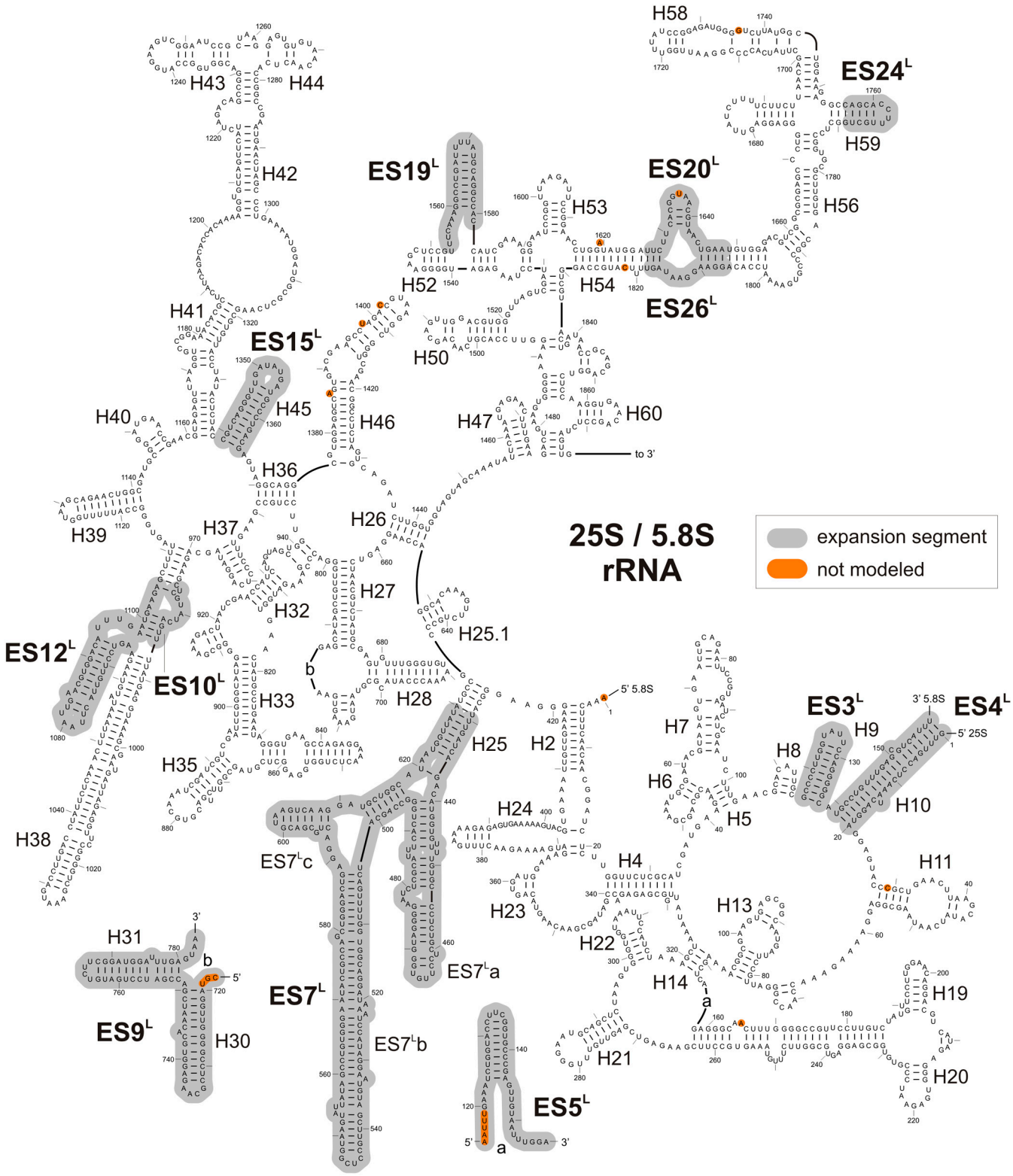


Fig. S6. Secondary structure diagram for the 5′ region of the large subunit rRNAs (5.8S and 25S) of *Saccharomyces cerevisiae* modified from ref. 1. Gray regions are expansion segments, whereas orange nucleotides were not modeled.

1 Cannone JJ, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

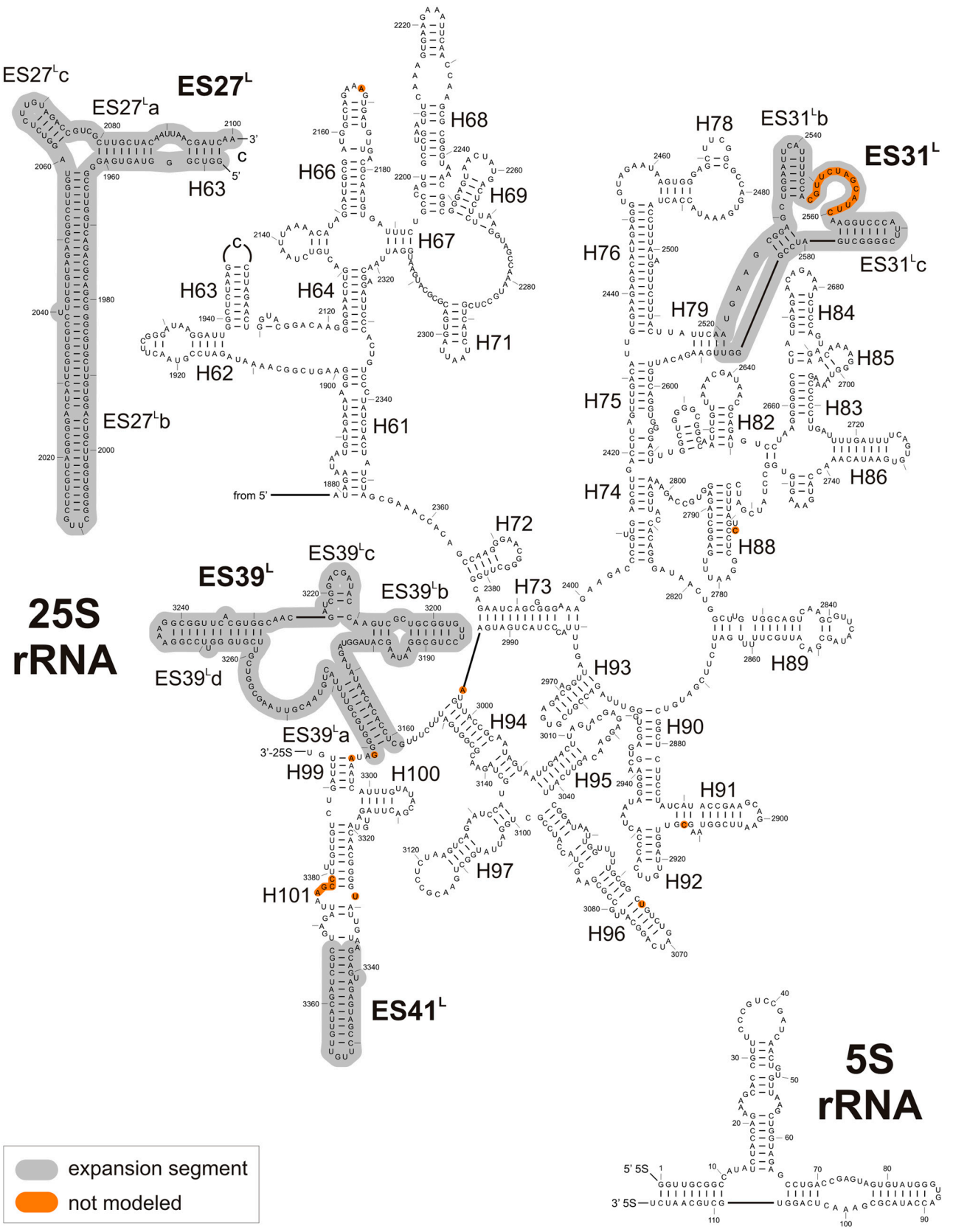


Fig. S7. Secondary structure diagram for the 3′ region of the large subunit rRNAs (5S and 25S) of *Saccharomyces cerevisiae* modified from ref. 1. Gray regions are expansion segments, whereas orange nucleotides were not modeled.

1 Cannone JJ, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

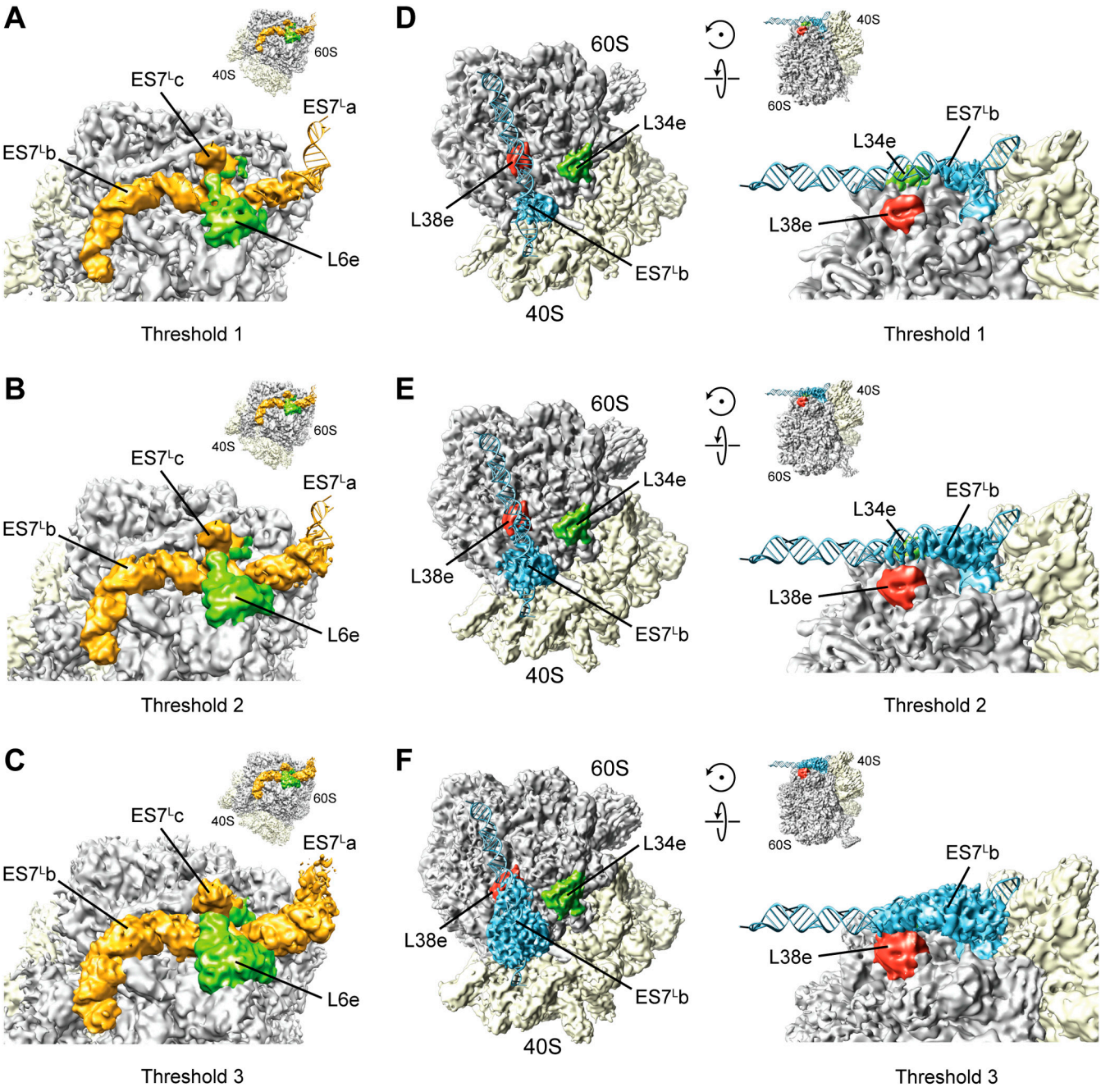


Fig. S8. Visualization of density for (*A*–*C*) *Saccharomyces cerevisiae* ES7^{L} (yellow) and (*D*–*F*) *Triticum aestivum* ES27$^{L}{}_{int}$ (blue) at different contour levels. (*A*–*C*) In the *S. cerevisiae* 80S reconstruction, density for ES7^{L}a is observed at lower thresholds (*C*). R-protein L6e (green) interacts with ES7^{L}b,c. (*D*–*F*) In the *T. aestivum* 80S reconstruction, density for ES27^{L}b is observed at lower thresholds, where interaction with r-protein L38e (red) with ES27^{L}b is evident (*F*).

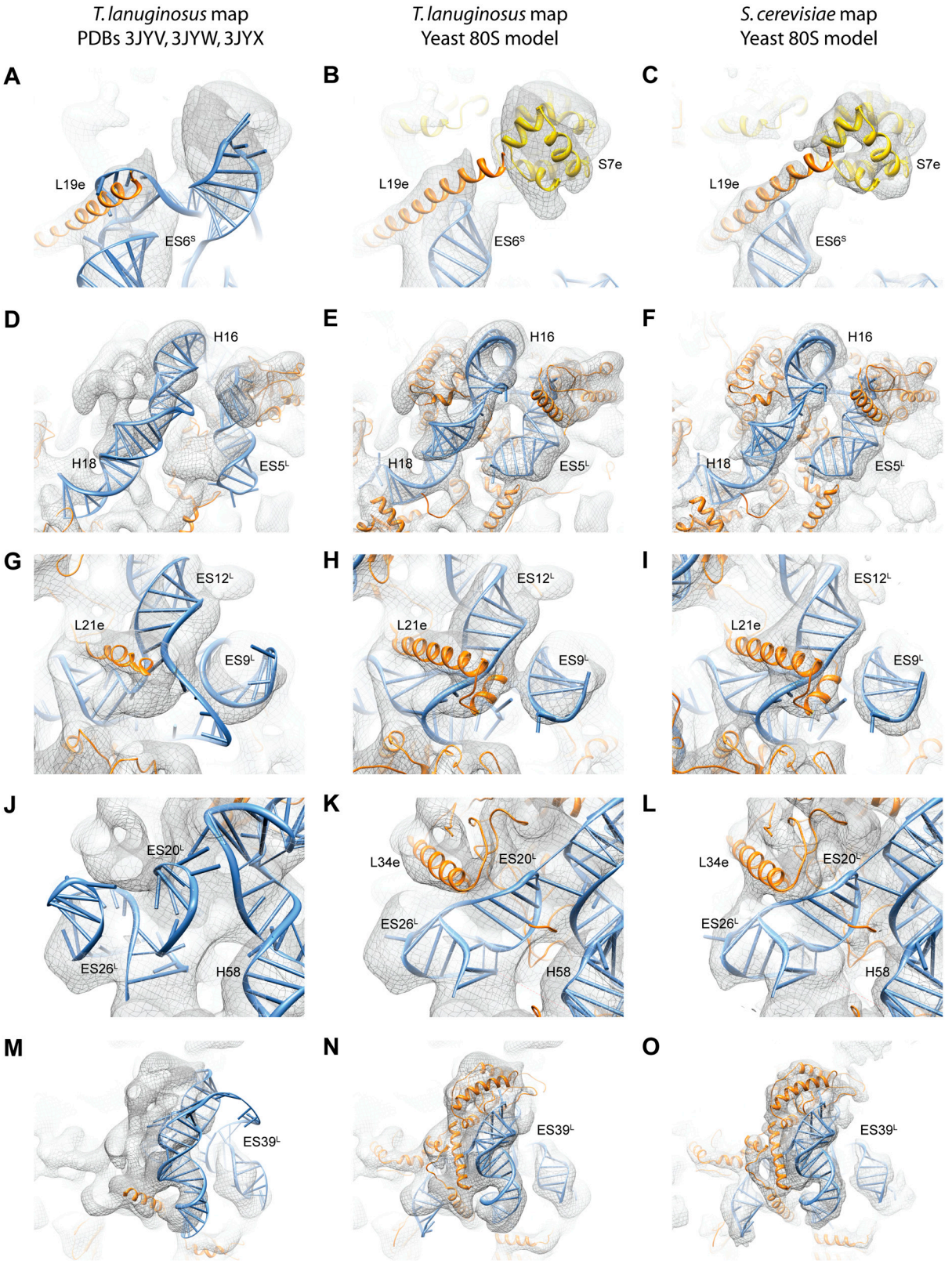


Fig. S9. Comparison of fit of yeast models in *Thermomyces lanuginosus* and *Saccharomyces cerevisiae* 80S cryo-EM maps. (*Left*) Protein Data Bank (PDB) 3JYV, 3JYW, 3JYX in the associated cryo-EM map of *T. lanuginosus* 80S ribosome from Taylor et al. (1). The fit of the yeast model (this paper) into cryo-EM map of *T. lanuginosus* 80S ribosome from Taylor et al. (1) and into the yeast 80S ribosome from Becker et al. (2) is shown in the center and right panels, respectively.

1 Taylor DJ, et al. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure* 17:1591–1604.
2 Becker T, et al. (2009) Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* 326:1369–1373.

Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome

Jean-Paul Armache[a,1], Alexander Jarasch[a,1], Andreas M. Anger[a,1], Elizabeth Villa[b], Thomas Becker[a], Shashi Bhushan[a], Fabrice Jossinet[c], Michael Habeck[d,e], Gülcin Dindar[a], Sibylle Franckenberg[a], Viter Marquez[a], Thorsten Mielke[f,g], Michael Thomm[h], Otto Berninghausen[a], Birgitta Beatrix[a], Johannes Söding[a], Eric Westhof[c], Daniel N. Wilson[a,2], and Roland Beckmann[a,2]

[a]Gene Center and Center for integrated Protein Science Munich, Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany; [b]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany; [c]Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique, 15 Rue René Descartes, 67084 Strasbourg, France; [d]Department of Empirical Inference, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany; [e]Department of Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany; [f]UltraStrukturNetzwerk, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany; [g]Institut für Medizinische Physik und Biophysik, Charité, Ziegelstrasse 5-8, 10117 Berlin, Germany; and [h]Universität Regensburg, Lehrstuhl für Mikrobiologie, Universitätstrasse 31, 93053 Regensburg, Germany

Edited* by Günter Blobel, The Rockefeller University, New York, NY, and approved September 17, 2010 (received for review July 10, 2010)

Protein synthesis in all living organisms occurs on ribonucleoprotein particles, called ribosomes. Despite the universality of this process, eukaryotic ribosomes are significantly larger in size than their bacterial counterparts due in part to the presence of 80 r proteins rather than 54 in bacteria. Using cryoelectron microscopy reconstructions of a translating plant (*Triticum aestivum*) 80S ribosome at 5.5-Å resolution, together with a 6.1-Å map of a translating *Saccharomyces cerevisiae* 80S ribosome, we have localized and modeled 74/80 (92.5%) of the ribosomal proteins, encompassing 12 archaeal/eukaryote-specific small subunit proteins as well as the complete complement of the ribosomal proteins of the eukaryotic large subunit. Near-complete atomic models of the 80S ribosome provide insights into the structure, function, and evolution of the eukaryotic translational apparatus.

homology modeling | RNA | translation | flexible fitting | molecular dynamics

Protein synthesis occurs on large macromolecular complexes, called ribosomes (1). Ribosomes are composed of two subunits, both of which are built from protein and RNA. Bacterial ribosomes, for example, in *Escherichia coli*, contain a small subunit composed of one 16S rRNA and 21 ribosomal proteins (r proteins), and a large subunit containing 5S and 23S rRNAs and 33 r proteins. In contrast, eukaryotic ribosomes are much larger and more complex, containing additional RNA in the form of so-called expansion segments (ES) as well as many additional r proteins and r-protein extensions. The additional r proteins present in eukaryotic ribosomes are likely to reflect the increased complexity of translation regulation in eukaryotic cells (2–5). Moreover, many of these eukaryote-specific components have been associated with human disorders (4). Thus, structural insight into the localization of these elements will be important to furthering our understanding of eukaryotic translation regulation as well as disease.

Compared with the ~54 r proteins of the bacterial ribosome, plant and fungal 80S ribosomes contain ~80 r proteins (see Table S1 for r-protein nomenclature). Crystal structures have revealed the location of each small and large subunit r protein within bacterial ribosomes (6–12) as well as the r proteins within the archaeal large ribosomal subunit (13, 14). In contrast, the localization of ribosomal proteins within eukaryotic 80S ribosomes has come mainly from early studies using immuno-EM and cross-linking approaches (see, for example, refs. 15–18). Moreover, the first molecular models for the eukaryotic ribosome were built at 15-Å resolution by docking the structures of the bacterial small 30S subunit (6) and archaeal large 50S subunit (13), thus only identifying the location of a total of 46 eukaryotic r proteins with bacterial or archaeal homologues (19). Recently, cryo-EM reconstructions of plant and fungal 80S ribosomes have led to the localization of three eukaryote-specific r proteins: RACK1 (20) and S19e (21) in the small subunit and L30e in the large subunit (22). Therefore, the current locations of 49 (33 large and 16 small subunit) r proteins are known for the eukaryotic 80S ribosome, whereas 31 (14 and 17, respectively) remain to be elucidated.

Here we have utilized cryo-EM maps of yeast and wheat germ ribosomes at 5.5 Å (see accompanying article in this issue of PNAS) and 6.1-Å resolution, respectively, to identify the location and build models for 74 of the 80 r proteins in the eukaryotic 80S ribosome, including 12 archaea/eukaryote-specific r proteins in the small subunit and 15 in the large subunit. Near-complete models for the yeast and wheat germ 80S ribosome will be an important resource for researchers working with these model organisms.

Results and Discussion

Placement of Ribosomal Proteins into a 5.5-Å Cryo-EM Map of an 80S Ribosome. Subtraction of the density assigned to the rRNA (gray in Fig. 1) in the 5.5-Å resolution cryo-EM structure of the *Triticum aestivum* translating 80S ribosome (see accompanying article in this issue of PNAS) left density that was attributed to r proteins (green in Fig. 1*A*). Due to the lack of complete sequence infor-

Author contributions: R.B. designed research; J.-P.A., A.J., A.M.A., E.V., T.B., S.B., F.J., G.D., S.F., V.M., T.M., O.B., B.B., E.W., and D.N.W. performed research; E.V., F.J., M.H., M.T., J.S., and E.W. contributed new reagents/analytic tools; J.-P.A., A.J., A.M.A., E.V., T.B., F.J., E.W., D.N.W., and R.B. analyzed data; and J.-P.A., A.J., A.M.A., D.N.W., and R.B. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: Coordinates of the atomic models of yeast and *Triticum aestivum* 80S complex have been deposited in the Protein Data Bank (PDB), www.pdb.org [PBD ID codes 3IZE, 3IZF, 3IZD (yeast rRNA), 3IZB, 3IZC (yeast r proteins), 3IZ7, 3IZ9 (*T. aestivum* rRNA), and 3IZ6, 3IZ5 (*T. aestivum* r proteins)]. The cryoelectron microscopic map of the *T. aestivum* 80S-RNCs has been deposited in the 3D-Electron Microscopy Data Bank (EMDB, http://www.ebi.ac.uk/pdbe/emdb/) (EMBD ID code EMD-1780).

[1]J.-P.A., A.J., and A.M.A. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: beckmann@lmb.uni-muenchen.de or wilson@lmb.uni-muenchen.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010005107/-/DCSupplemental.

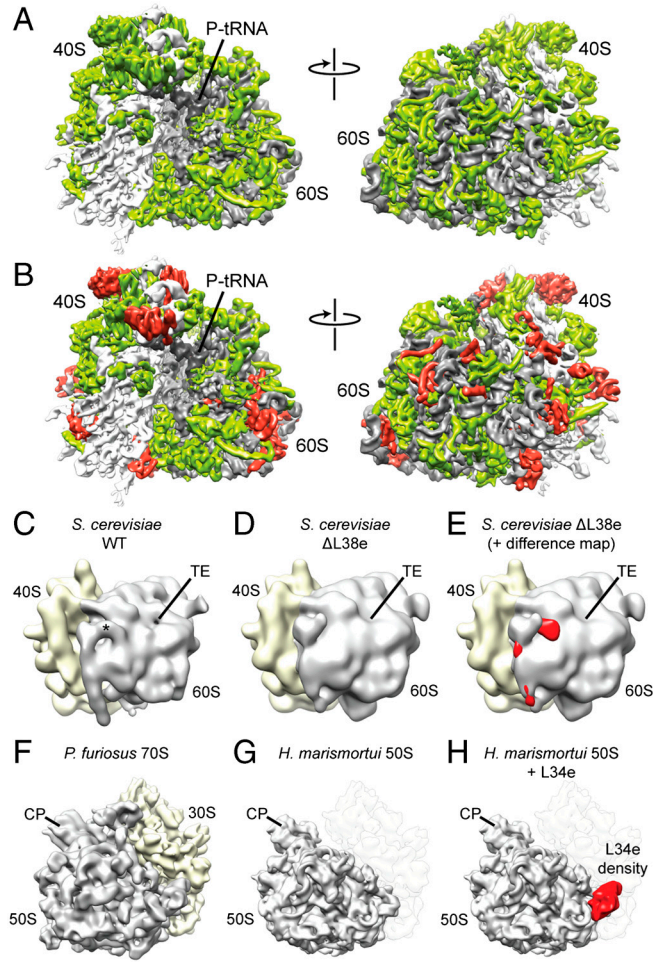


Fig. 1. Identification of r-proteins L38e and L34e. (*A*) Cryo-EM map of the *T. aestivum* 80S ribosome, with rRNA colored gray and r protein colored green. (*B*) Same as *A*, but with localized r proteins colored red. Reconstruction of (*C*) *S. cerevisiae* WT 80S ribosome compared to (*D*) reconstruction of *S. cerevisiae* 80S ribosomes isolated from a strain lacking the gene for L38e. The asterisk indicates the position of additional density assigned to L38e, and the tunnel exit (TE) is shown for reference. (*E*) Difference density map calculated between *C* and *D* and shown superimposed on the map from *D*. Reconstruction of (*F*) *P. furiosus* 70S ribosome, compared to (*G*) X-ray structure of the 50S subunit from *H. marismortui* filtered to a similar resolution. (*H*) Difference density map calculated between *F* and *G* and shown superimposed on the map from *G* identifying the location of r-protein L34e (red).

mation for *T. aestivum*, sequences of the closely related *Oryza sativa* were used where necessary (Tables S1–S6). This is a valid approach because of the given resolution of the map and the very high similarity of the proteins (>90% identity on average). Models for 44 of the 80 r proteins of the *T. aestivum* 80S ribosome were built into this map using the templates present in the bacterial and archaeal ribosome structures (23, 24). Similarly, 44 of 79 r proteins of the yeast 80S ribosome were built into the previously reported cryo-EM structure of a translating *Saccharomyces cerevisiae* 80S ribosome at 6.1-Å resolution (25). The archaeal/eukaryote-specific r-protein extensions were modeled de novo whenever possible, building out from N and C terminus of the template-based core regions using electron density map and secondary structure prediction constraints.

A total of 17 r proteins (see Table S1 for r-protein family nomenclature), 7 (S4e, S17e, S19e, S24e, S27e, S28e, and RACK1) from the 40S subunit, and 10 (L4e, L6e, L14e, L18ae, L27e, L30e, L35ae, P0, P1, and P2) from the 60S subunit were modeled using available X-ray and NMR structures of free r proteins (Tables S2–S5). Homology models for six r proteins (S25e, L22e, L29e, L34e, L36e, and L38e) were built using HHpred (26) and Modeller (27) on the basis of similarity with domains of proteins of known structure, for example, S25e and L38e were predicted to have helix-turn-helix and K-homology domains, both of which are known to interact with RNA. Seven r proteins (S7e, S21e, S26e, S30e, L13e, L28e, and L41e) were tentatively modeled ab initio on the basis of secondary structure predictions and density characteristics, and six small subunit r proteins (S3ae, S6e, S8e, S10e, S12e, and S27a) could not be localized and were therefore not modeled. The protein models were initially fitted as rigid bodies, merged with the rRNA models and an extended version of molecular dynamics flexible fitting (MDFF) was applied to remove clashes, impose stereochemical restraints, and improve the overall fit to the map (28). At the given resolution, it should be noted that the degree of accuracy and reliability of the assignments varies for the different r proteins (Table S6): The fold and location of ribosomal core proteins and those modeled on the basis of available X-ray and NMR structures will have a higher degree of accuracy than those generated using remote homology or ab initio modeling. Although the latter models can only be considered tentative placements, the location of the r protein is more certain, being consistent with available biochemical evidence (Table S6). Collectively, a total of 74 r proteins were modeled, 27 (excluding P0, P1, P2) of which are not present in the bacterial or archaeal ribosome crystal structures (red in Fig. 1*B*).

Localization of Ribosomal Proteins of the 80S Ribosome. The main basis for the localization of r proteins in the cryo-EM reconstructions of the 80S ribosomes was the excellent agreement between the density features in the maps with distinctive protein-fold characteristics of the X-ray structures and homology models (Fig. S1). Additional supporting information was utilized for the localization of r proteins, particular those modeled ab initio. The supporting data included species-specific differences in length between r proteins of wheat germ, yeast, and archaeal ribosomes, as well as the wealth of data available on the spatial arrangement of r proteins in eukaryotic ribosomes derived from a variety of different approaches: (*i*) the order of assembly of r proteins (29); (*ii*) accessibility of particular r proteins to proteolysis; (*iii*) cross-linking of r proteins (15, 18, 30, 31); and (*iv*) immuno-EM studies (16, 32) (see Table S6). Furthermore, the localization of r-protein L38e was supported by comparison of a cryo-EM reconstruction of wild-type yeast 80S ribosome with that of a yeast 80S ribosome isolated from a strain lacking the gene for r-protein L38e (Fig. 1 *C–E*). Similarly, comparison of a cryo-EM reconstruction of a 70S ribosome from the archaeon *Pyrococcus furiosus* with the crystal structure of the large subunit of *Haloarcula marismortui* led to the localization of r-protein L34e (Fig. 1 *F–H*). Both L34e and L38e stabilize different conformations of ES27 (see accompanying article in this issue of PNAS). R-protein L28e is not present in the *S. cerevisiae* genome, and therefore the localization of L28e was possible by generating difference maps between yeast and *T. aestivum* ribosomes (Fig. S2).

On this basis, it was possible to localize and model a total of 27 r proteins (excluding P0, P1, and P2) that are not present in the crystal structures of bacterial or archaeal ribosomes (Fig. 2 *A–D*). This encompasses 12 small subunit r proteins (S4e, S7e, S17e, S19e, S21e, S24e, S25e, S26e, S27e, S28e, S30e, and RACK1) and 15 large subunit r proteins (L6e, L13e, L14e, L18ae, L22e, L27e, L28e, L29e, L30e, L34e, L35ae, L36e, L38e, L40e, and L41e) (Fig. 2 *A–D*). We can assign the unidentified protein interaction partner of RACK1 (33) as being the eukaryote-specific C-terminal extension of r-protein S2p, whereas the localization of L30e on the 60S subunit is as reported for yeast and *T. aestivum* (Fig. 2 *B* and *D*) (22, 34). Mutations in S19e found in Diamond-Blackfan anemia (DBA) patients are clustered around α3 (35), which is seen to interact with h41 in the *T. aestivum* and yeast

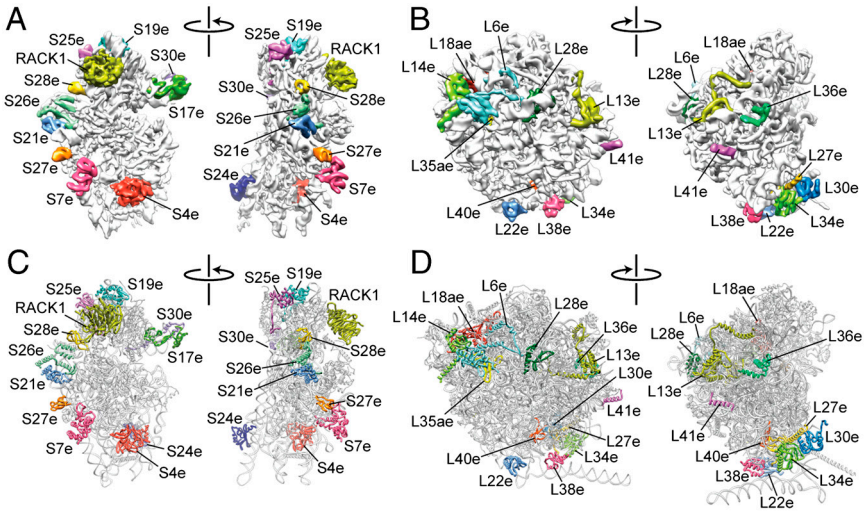


Fig. 2. Localization of eukaryote-specific r proteins. Cryo-EM maps of the *T. aestivum* (*A*) 40S and (*B*) 60S subunit, with density for the newly identified r proteins colored distinctly. Molecular models of r proteins of the *T. aestivum* (*C*) 40S and (*D*) 60S subunit, with newly identified r proteins colored distinctly.

80S models. DBA is an inherited bone marrow failure syndrome that results from defects in ribosomal assembly (4). The localization of S19e (and S28e) to the head of the 40S subunit (Fig. 2*A*) is also consistent with biochemical data examining assembly precursor particles formed in vivo (29). In addition to S19e, we have localized the other major r proteins associated with DBA, such as S7e on the platform at the base of ES6, S17e to the beak of the 40S subunit, as well as S24e at the interface side bridging h8 and h44 (Fig. 2*A*).

Functional Roles for Eukaryote-Specific Ribosomal Proteins. Although the active sites of the ribosome—the decoding site on the small subunit and the site of peptide-bond formation on the large subunit—are composed largely of rRNA, they are not completely devoid of r proteins (Fig. 3 *A–D*). Compared with bacterial 30S subunits, eukaryotic 40S subunits contain two additional r proteins, S25e and S30e, with extensions that reach into the decoding and tRNA binding sites (Fig. 3 *A* and *B*). Consistent with this localization, S30e has been cross-linked to the 4-thiouridine containing UGA stop codon of mRNA positioned at the A-site (30). Additionally, the C terminus of r-protein S4p is relocated in eukaryotes, due to corresponding rearrangements in h16/17, and reaches from the globular domain on the solvent side right into the decoding site of the small subunit (Fig. 3*A*). Thus,

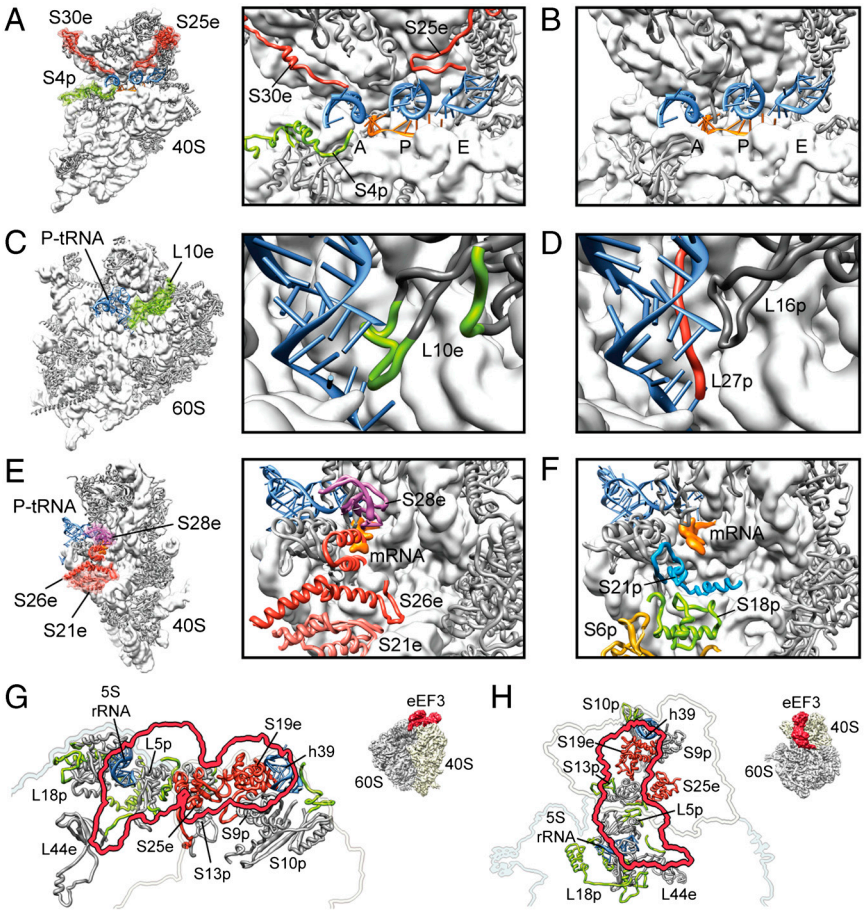


Fig. 3. Functional roles for eukaryote-specific r proteins. (*A*) Small 40S subunit with newly modeled r-proteins S30e and S25e (red) and eukaryote-specific extension of S4p (green) highlighted (thumbnail, *Left*; zoom, *Right*). (*B*) Comparative view of the bacterial 30S subunit decoding site (11, 12). In *A* and *B*, the anticodon-stemloops of A-, P- and E-tRNAs (blue) and mRNA (orange) are shown for reference. (*C*) Large 60S subunit with eukaryote-specific extension of L10e (green) highlighted (thumbnail, *Left*; zoom, *Right*). (*D*) Comparative view of the bacterial 50S subunit with bacterial-specific L27p colored red (11). In *C* and *D*, the acceptor-stem of the P-tRNA (blue) is shown for reference. (*E*) Small 40S subunit with newly modeled r-proteins S21e, S26e, and S28e colored distinctly (thumbnail, *Left*; zoom, *Right*). (*F*) Comparative view of the bacterial 30S subunit with bacterial-specific S18p shown in green (11). In *E* and *F*, the P-tRNA (blue) and mRNA (orange) are shown for reference. (*G* and *H*) The binding site of eEF3 on the *S. cerevisiae* 80S ribosome, with (*G*) side and (*H*) top views (see insets) showing the binding site of eEF3 as a red outline and molecular models of ribosomal components that comprise the eEF3 binding site. Newly identified proteins are shown in red (S19e, S25e) and newly modeled r-protein extensions in green, whereas core r proteins are colored gray. Modified from ref. 48.

together with the extensions and loops of eukaryotic homologues to the bacterial S7, S9, S11, S12, and S13 r proteins (11, 12), at least seven different r proteins can interact and modulate the binding of tRNAs to the 40S subunit. At the peptidyl-transferase center on the large subunit, direct interaction is observed between the loop of r-protein L10e and the CCA-end of a peptidyl-tRNA at the P site (Fig. 3*C*). Based on our model, the loop of L10e is now the r-protein region that comes closest (~16 Å) to the site of peptide-bond formation (Fig. S3). This loop was disordered and not modeled in the crystal structures of the archaeal 50S subunit (13) and is absent in the bacterial homologue, L16. Instead, the N-terminal extension of the r-protein L27 occupies a similar but distinct position in bacterial ribosomes (36, 37) (Fig. 3*D*). The loop of L10e is highly conserved and mutations or deletions in this loop are lethal (38), suggesting that it may play an important role in tRNA positioning, as proposed for the N terminus of L27 (36, 37).

Three eukaryote-specific r proteins, S21e, S26e, and S28e, were identified at the mRNA exit site between the platform and head of 40S subunit (Fig. 3*E*). Both S26e and S28e have been cross-linked from positions (−6 and −7/ − 10, respectively) in the 5′ untranslated region (UTR) of mRNA (18). The equivalent region of bacterial 30S subunits is occupied by bacterial-specific r proteins S6, S8 as well as S21 in *E. coli* (6, 10) (Fig. 3*F*). These differences may reflect the distinct elements found in the 5′ UTRs of bacterial and eukaryotic mRNAs, as well as the divergence in the translation initiation phase (2). For example, eIF3, which is absent in bacteria, interacts with this region of the eukaryotic 40S subunit (32, 39–41). Internal ribosome entry site (IRES) elements present in the 5′ UTR of viral mRNAs also interact with this region of eukaryotic ribosomes (42–45). Indeed, the unknown rpSx that interacts with cricket paralysis virus (CrPV) IRES (45) can now be assigned as r-protein S25e, consistent with a cross-link from the conserved domain 2 fragment from CrPV and other IRESs to S25e (31).

The translation factor binding site is highly conserved on bacterial and eukaryotic ribosomes, with the exception of the aforementioned extensions of r proteins S4p and S30e that reach into the decoding site in the 80S ribosome. Extensions of both S4p and S30e would be expected to interact with domain IV of eEF2, as visualized previously by cryo-EM (46, 47). Additionally, we can now identify the eEF3 interaction partners in the yeast 80S, previously assigned as rpSX1 and rpSX2 (48), as being r-protein S19e and S25e, respectively, both of which are located in the head of the 40S subunit (Fig. 3 *G* and *H*). In addition, r-protein L44e as well as eukaryote-specific extensions of r proteins L5p and L18p located within the central protuberance of the 60S subunit also comprise the eEF3 binding site (Fig. 3 *G* and *H*).

Coevolution of rRNA Expansion Segments and Eukaryotic-Specific Ribosomal Proteins. Eukaryotic 80S ribosomes are significantly larger than their bacterial counterparts, the *T. aestivum* ribosome contains 1.53 MDa (0.62 MDa/40S and 0.91 MDa/60S) of r protein and 1.74 MDa (0.56 MDa/40S and 1.18 MDa/60S) of rRNA, thus totaling 3.27 MDa, whereas *E. coli* 70S ribosomes total to ~2.5 MDa (0.9 MDa/30S and 1.6 MDa/50S). Fig. 4*A* shows that the ES and additional r proteins/protein extensions (green and gold, respectively) form an intricate layer of additional RNA-protein mass that locates predominantly to the solvent surface of the ribosome. The intertwined nature of the additional rRNA ES and r proteins supports the idea that they are coevolving together (49), which is exemplified by the large mass found on the back of the 60S subunit comprising ES7^{L}, ES39^{L}, and five eukaryotic r proteins (L6e, L14e, L18ae, L28e, and L35ae) (Fig. 4*B*). Interestingly, L6e, L14e, and L27e all adopt the same SH3-like barrel fold, possibly reflecting their origin due to gene duplication events. L27e is located below the L1 stalk on the opposite side of the ribosome from L6e and L14e, where it is sandwiched between

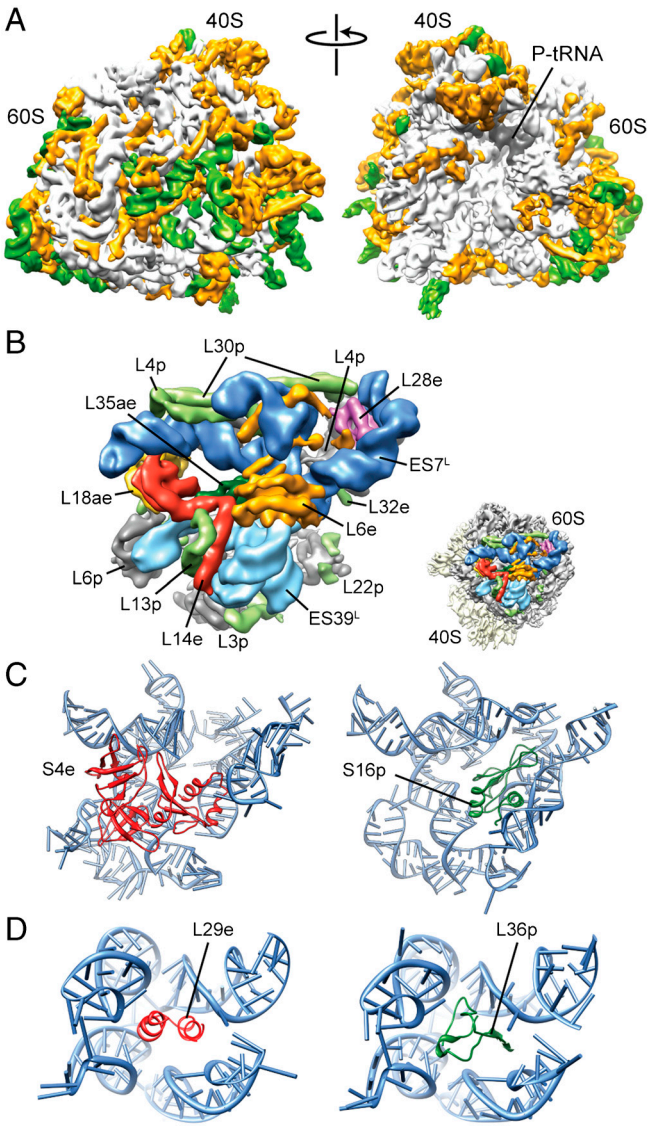


Fig. 4. Coevolution of rRNA expansion segments with r proteins in the 80S ribosome. (*A*) Cryo-EM map of the *T. aestivum* 80S ribosome, with rRNA ES and variable regions colored green and eukaryote-specific r proteins and extensions colored orange. (*B*) View of the intertwined region of ES7^{L} (dark blue) and ES39^{L} (light blue), with core r proteins (gray), eukaryote-specific r-protein extensions (pale green), and r proteins (L6e, orange; L14e, red; L18ae, yellow; L28e, pink; L35ae, green) highlighted. Inset shows relative position to 40S (yellow) and 60S subunits (gray). (*C*) Comparison of relative positions of S4e (red) in yeast/*T. aestivum* 80S (*Left*) with S16p (green) in bacteria (11) (*Right*). (*D*) Comparison of relative positions of L29e (red) in yeast/*T. aestivum* 80S (*Left*) with L36p (green) in bacteria (11) (*Right*).

H55 and H58. L27e and L34e overlap the position of H58 in the *E. coli* 70S ribosome, emphasizing the conformational rearrangements that relocate H58 in archaeal/eukaryotic compared to bacterial ribosomes. In contrast, r proteins, such as L13e, L22e, and L36e, occupy empty sites in the bacterial and archaeal ribosomes yet interact with the core rRNA. Interestingly, the loop of H57, which is the interaction partner for L22e, is conserved in eukaryotes, but not in bacteria, which lack this protein.

Evolution of the Eukaryotic Ribosome. A previous comparison of archaeal and bacterial large subunits illustrated examples of potential convergent evolution, where evolutionarily unrelated r proteins have evolved to stabilize the same region of 23S rRNA (14). Many such examples are also found by comparing the mod-

els of the yeast and *T. aestivum* 80S ribosome with the archaeal and bacterial crystal structures: The N-terminal domain of S4e overlaps the binding position of S16p (Fig. 4*C*), and the extended N terminus of L32e overlaps regions of bacterial-specific r proteins L20p and L21p. Likewise, L18ae has two ubiquitin-like α/β roll domains (ULDs), with the N-terminal ULD overlapping bacterial L25p, and like L25p also interacting with the 5S rRNA, whereas α-helix 1 of the C-terminal ULD inserts in the minor groove of H41. Furthermore, L29e sits in a small RNA pocket at the stalk base, which is occupied by L36p in bacteria (Fig. 4*D*). The localization of L29e to this pocket was based partly on the observation that the stalk rearranges position to establish contact with the head of the 40S subunit in a reconstruction of the yeast ΔL29e-80S (Fig. S4), which has not been observed in any previous yeast 80S reconstructions. Moreover, the assigned position for L29e is in close proximity to L10e (L16p), which exhibits synthetic lethality with L29e in yeast (50).

Conclusion

Molecular models are presented for translating *T. aestivum* and yeast 80S ribosomes encompassing ~98% of the rRNA and 92.5% of the r proteins (Fig. 5). Given that mammalian ribosomes have the same complement of 80 r proteins as those of *T. aestivum* presented here, we believe that the information gained from the *T. aestivum* and yeast 80S models should thus not only provide a resource for researchers working with these model organisms, but may also provide useful information when studying mammalian systems.

Experimental Procedures

Sample Preparation and Cryoelectron Microscopy. Yeast (*Saccharomyces cerevisiae*) 80S ribosomes were isolated from wild-type strains and strains lacking the genes encoding r proteins L29e and L38e (29), as described previously (25, 51). Cryo-EM reconstructions of the yeast ΔL29e and ΔL38e-80S ribosomes were performed on a Tecnai G2 Spirit transmission electron microscope at 120 kV at a nominal magnification of 90,000 using an Eagle 4,096 × 4,096 pixel CCD camera (FEI) resulting in a pixel size of 3.62 Å/pixel. For the final yeast ΔL29e and ΔL38e-80S ribosome reconstructions, 7,272 and 10,356 particles were used. The cryo-EM map used for modeling of the yeast 80S ribosome was published previously (Electron Microscopy Data Bank ID 1669; ref. 25). Cryo-EM reconstructions of the *P. furiosus* 70S ribosomes were performed as for the *T. aestivum* ribosome nascent chain complex samples described in the accompanying article in this issue of PNAS. The final reconstruction of the *P. furiosus* 70S ribosome used 54,979 particles, yielding a final contrast transfer function corrected map at a resolution of 10 Å. Densities for the 40S subunit, the 60S subunit, and the P-site tRNA were isolated using binary masks.

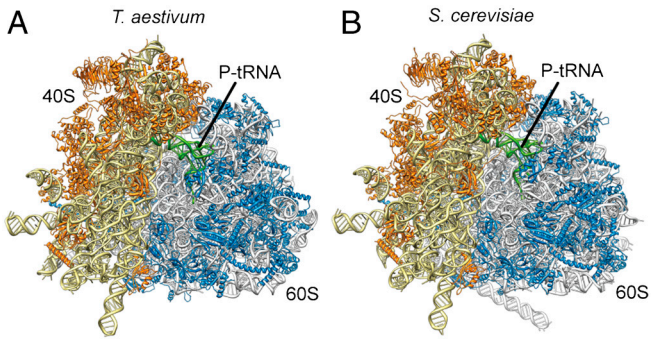


Fig. 5. Structures of wheat germ and yeast eukaryotic 80S ribosomes. (*A* and *B*) Near-complete molecular models for the (*A*) *T. aestivum* and (*B*) *S. cerevisiae* 80S ribosome, with rRNA and protein shown in yellow and orange for the small subunit and gray and blue for the large subunit, respectively.

Density Map Sharpening. For modeling of r-protein extensions, density maps were sharpened using a nonnegative deconvolution method (Hirsch, Schölkopf and Habeck, accepted) based on the multiplicative updates proposed in (52). As a blurring function, an isotropic Gaussian kernel (generated with the EMAN software package command *pdb2mrc* for a Protein Data Bank file containing a single atom) was chosen. In addition, a nonnegative background density was introduced to account for solvent contributions and other artefacts. The background was constrained to be uncorrelated with the deconvolved density map. Both the deconvolved map and the background density were then estimated simultaneously using interleaved multiplicative updates. The deconvolution algorithm was run for different kernel sizes and constraint strengths. The most informative density map was selected by visual inspection (Fig. S1).

Homology Modeling of R Proteins. Based on the crystal structures of the archaeal 50S subunit (13) and the bacterial ribosomal structures (10, 11), it was possible to generate *S. cerevisiae* and *T. aestivum* (or *O. sativa*) homology models (Tables S2–S5). In addition, there are also 12 structures of r proteins obtained from either X-ray or NMR structures in a non-ribosome-associated state (Tables S2–S5). The best templates were chosen by screening available structures and selecting on the basis of both the sequence identity and fitting to the cryo-EM density. Sequence to structure matching has been performed based on profile–profile alignments (53, 54). Alignments were performed using a number of alignment servers, including ClustalW (55), TCoffee (56), MUSCLE (57), and Mafft (58). Using Modeler (59), numerous models were created and ranked based on the discrete optimized protein energy (60) score. From the top scoring models, two were chosen and rigidly fitted into the EM density using Chimera (61) and Coot (62), and the best fit was taken for further refinement. Extended parts of the proteins that did not have a template were truncated at this step and manual adjustments were introduced to the rigidly fitted protein to best fit the density. Whenever modeling of extensions appeared possible on the basis of information in the cryo-EM map, secondary structure predictions were performed (63), along with search for an appropriate template among existing structures using HHpred (64). This information, together with the density information in close proximity to the protein core was examined and, if possible, the extended part was modeled. In cases of ambiguous density, comparison of additional maps (*S. cerevisiae*, deconvolved *T. aestivum*; see Fig. S1) was used. Using this approach, a total of over 2,000 amino acids were modeled de novo. The increasing number of modeled extensions allowed us to iteratively minimize the amount of available density, thus providing constraints to find additional solutions to RNA and protein localization.

Refinement and Fitting of the R Proteins into the EM Densities. Because common methods for protein modeling are to date not capable of incorporating EM data or interaction with RNA directly in the modeling process, the proteins still needed to be flexibly fitted into the density and reconciled with RNA models. Thus, subsequent to the fitting and modeling of the rRNA, proteins were introduced in the model using visual molecular dynamics (VMD) (65), and interactive MDFF was used to refine the proteins into the density using default parameters (28). In regions where the protein density was weak, the location of protein regions was determined by visual inspection, and harmonic constraints to the alpha carbons of those regions were imposed to preserve such location. This process resulted in a rearrangement of the proteins to fit the density, and to resolve protein–RNA and protein–protein clashes while preserving secondary structure. Further MDFF refinement was then performed on the entire 80S model. The fitting was performed iteratively, starting with the most reliable fits, such as docking of X-ray structures and

homology models of r proteins. This reduced the leftover density available for localization and modeling of the remaining unassigned r proteins or r-proteins extensions, that later underwent further refinement.

Visualization and Figure Preparation. Cryo-EM maps and models were visualized and all figures were generated using VMD (65), Chimera (61), and/or PyMol (http://www.pymol.org).

ACKNOWLEDGMENTS. This research was supported by grants from the Deutsche Forschungsgemeinschaft SFB594 and SFB646 (to R.B.), SFB740 (to T.M.), and WI3285/1-1 (to D.N.W.), by the European Union and Senatsverwaltung für Wissenschaft, Forschung und Kultur Berlin (UltraStructure-Network, Anwenderzentrum), and a Marie Curie International Incoming Fellowship within the seventh European Community Framework Programme (E.V.). Computer time for MDFF was provided by the Leibniz-Rechenzentrum.

1. Schmeing TM, Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461:1234–1242.
2. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: Mechanisms and biological targets. *Cell* 136:731–745.
3. Warner JR, McIntosh KB (2009) How common are extraribosomal functions of ribosomal proteins? *Mol Cell* 34:3–11.
4. Freed EF, Bleichert F, Dutca LM, Baserga SJ (2010) When ribosomes go bad: Diseases of ribosome biogenesis. *Mol Biosyst* 6:481–493.
5. Wang DO, Martin KC, Zukin RS (2010) Spatially restricting gene expression by local translation at synapses. *Trends Neurosci* 33:173–182.
6. Wimberly BT, et al. (2000) Structure of the 30S ribosomal subunit. *Nature* 407:327–339.
7. Schluenzen F, et al. (2000) Structure of functionally activated small ribosomal subunit at 3.3 A resolution. *Cell* 102:615–623.
8. Yusupov MM, et al. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* 292:883–896.
9. Harms J, et al. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* 107:679–688.
10. Schuwirth B, et al. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310:827–834.
11. Selmer M, et al. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313:1935–1942.
12. Jenner LB, Demeshkina N, Yusupova G, Yusupov M (2010) Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol* 17:555–560.
13. Ban N, et al. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289:905–920.
14. Klein D, Moore P, Steitz T (2004) The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* 340:141–177.
15. Gross B, Westermann P, Bielka H (1983) Spatial arrangement of proteins within the small subunit of rat liver ribosomes studied by cross-linking. *EMBO J* 2:255–260.
16. Marion MJ, Marion C (1987) Localization of ribosomal proteins on the surface of mammalian 60S ribosomal subunits by means of immobilized enzymes. Correlation with chemical cross-linking data. *Biochem Biophys Res Commun* 149:1077–1083.
17. Lutsch G, et al. (1990) Immunoelectron microscopic studies on the location of ribosomal proteins on the surface of the 40S ribosomal subunit from rat liver. *Eur J Cell Biol* 51:140–150.
18. Pisarev AV, et al. (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* 27:1609–1621.
19. Spahn CM, et al. (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions. *Cell* 107:373–386.
20. Sengupta J, et al. (2004) Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM. *Nat Struct Mol Biol* 11:957–962.
21. Taylor DJ, et al. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure* 17:1591–1604.
22. Halic M, et al. (2005) Localization and dynamic behavior of ribosomal protein L30e. *Nat Struct Mol Biol* 12:467–468.
23. Ban N, et al. (1999) Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature* 400:841–847.
24. Clemons WMJ, et al. (1999) Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* 400:833–840.
25. Becker T, et al. (2009) Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* 326:1369–1373.
26. Hildebrand A, Remmert M, Biegert A, Soding J (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins* 77(Suppl 9):128–132.
27. Eswar N, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* 5.6.1–5.6.30.
28. Trabuco LG, et al. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16:673–683.
29. Ferreira-Cerca S, et al. (2007) Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins. *Mol Cell* 28:446–457.
30. Bulygin K, et al. (2005) The first position of a codon placed in the A site of the human 80S ribosome contacts nucleotide C1696 of the 18S rRNA as well as proteins S2, S3, S3a, S30, and S15. *Biochemistry* 44:2153–2162.
31. Nishiyama T, Yamamoto H, Uchiumi T, Nakashima N (2007) Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. *Nucleic Acids Res* 35:1514–1521.
32. Bommer UA, Lutsch G, Stahl J, Bielka H (1991) Eukaryotic initiation factors eIF-2 and eIF-3: Interactions, structure and localization in ribosomal initiation complexes. *Biochimie* 73:1007–1019.
33. Chandramouli P, et al. (2008) Structure of the mammalian 80S ribosome at 8.7 A resolution. *Structure* 16:535–548.
34. Chao JA, Williamson JR (2004) Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. *Structure* 12:1165–1176.
35. Gregory LA, et al. (2007) Molecular basis of Diamond-Blackfan anemia: Structure and function analysis of RPS19. *Nucleic Acids Res* 35:5913–5921.
36. Maguire BA, et al. (2005) A protein component at the heart of an RNA machine: The importance of protein l27 for the function of the bacterial ribosome. *Mol Cell* 20:427–435.
37. Voorhees RM, et al. (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* 16:528–533.
38. Hofer A, Bussiere C, Johnson AW (2007) Mutational analysis of the ribosomal protein Rpl10 from yeast. *J Biol Chem* 282:32630–32639.
39. Westermann P, Nygard O (1983) The spatial arrangement of the complex between eukaryotic initiation factor eIF-3 and 40S ribosomal subunit. Cross-linking between factor and ribosomal proteins. *Biochim Biophys Acta* 741:103–108.
40. Westermann P, Nygaard O (1984) Crosslinking of mRNA to initiation factor eIF-3, 24 kDa cap binding protein and ribosomal proteins S1, S3/3a, S6 and S11 within the 48S pre-initiation complex. *Nucleic Acids Res* 12:8887–8897.
41. Srivastava S, Verschoor A, Frank J (1992) Eukaryotic initiation factor-3 does not prevent association through physical blockage of the ribosomal subunit-subunit interface. *J Mol Biol* 226:301–304.
42. Spahn CM, et al. (2001) Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* 291:1959–1962.
43. Spahn CM, et al. (2004) Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes; the IRES functions as an RNA-based translation factor. *Cell* 118:465–475.
44. Boehringer D, et al. (2005) Structure of the hepatitis C virus IRES bound to the human 80S ribosome: Remodeling of the HCV IRES. *Structure* 13:1695–1706.
45. Schuler M, et al. (2006) Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat Struct Mol Biol* 13:1092–1096.
46. Gomez-Lorenzo MG, et al. (2000) Three-dimensional cryo-electron microscopy localization of EF2 in the *Saccharomyces cerevisiae* 80S ribosome at 17.5 angstrom resolution. *EMBO J* 19:2710–2718.
47. Spahn CM, et al. (2004) Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J* 23:1008–1019.
48. Andersen BF, et al. (2006) Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature* 433:663–668.
49. Yokoyama T, Suzuki T (2008) Ribosomal RNAs are tolerant toward genetic insertions: Evolutionary origin of the expansion segments. *Nucleic Acids Res* 36:3539–3551.
50. DeLabre ML, Kessl J, Karamanou S, Trumpower BL (2002) RPL29 codes for a non-essential protein of the 60S ribosomal subunit in *Saccharomyces cerevisiae* and exhibits synthetic lethality with mutations in genes for proteins required for subunit coupling. *Biochim Biophys Acta* 1574:255–261.
51. Beckmann R, et al. (1997) Alignment of conduits for the nascent polypeptide chain in the ribosome- Sec61 complex. *Science* 278:2123–2126.
52. Sha F, Lin Y, Saul LK, Lee DD (2007) Multiplicative updates for nonnegative quadratic programming. *Neural Comput* 19:2004–2031.
53. Eswar N, Madhusudhan MS, Marti-Renom MA, Sali A (2005) PROFILE_SCAN: A module for fold assignment using profile-profile scanning in MODELLER. http://www.salilab.org/modeller.
54. Marti-Renom MA, Madhusudhan MS, Sali A (2004) Alignment of protein sequences by their profiles. *Protein Sci* 13:1071–1087.
55. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
56. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217.
57. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
58. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066.
59. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.
60. Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524.
61. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
62. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr, Sect D: Biol Crystallogr* 60:2126–2132.
63. Biegert A, et al. (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* 34:W335–339.
64. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–248.
65. Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. *J Mol Graphics* 14:33–38.

Supporting Information

Armache et al. 10.1073/pnas.1010005107

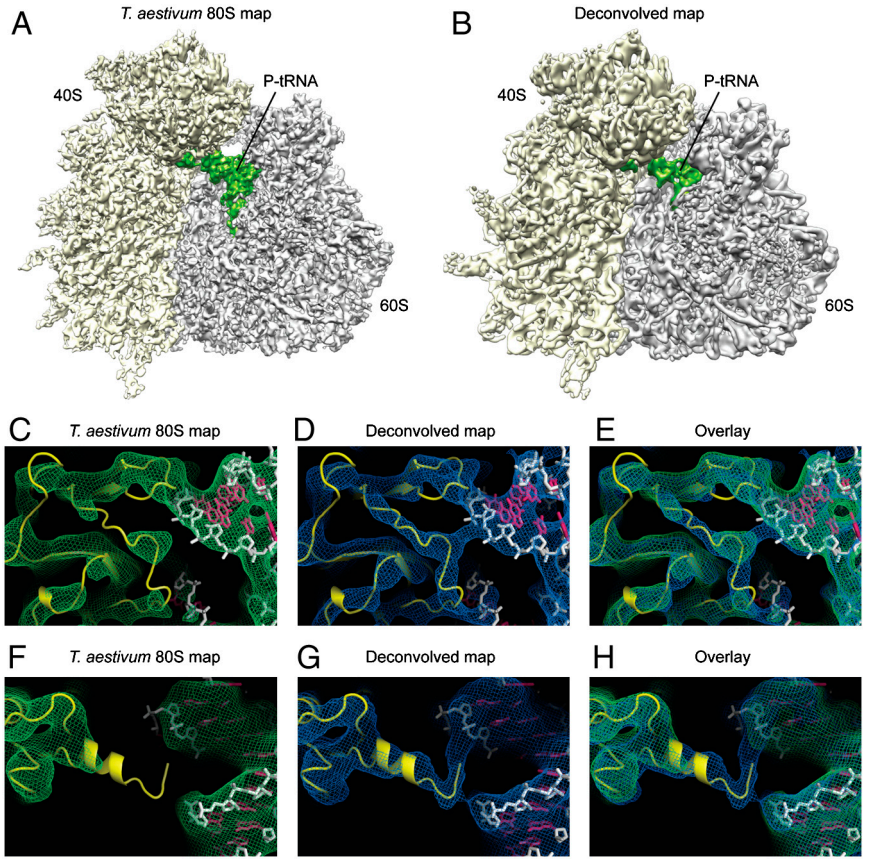


Fig. S1. Comparison of pre- and postdeconvolution sharpened maps. (*A*) pre- and (*B*) postdeconvolution cryo-EM maps of the *Triticum aestivum* 80S ribosome, with small and large subunits in yellow and gray, respectively, and P-tRNA colored green. Examples of (*C* and *F*) pre- and (*D* and *G*) postdeconvolution maps (mesh), with overlays shown in *E* and *H*, respectively. R proteins are shown as yellow ribbons and rRNA nucleotides with white backbone and red bases.

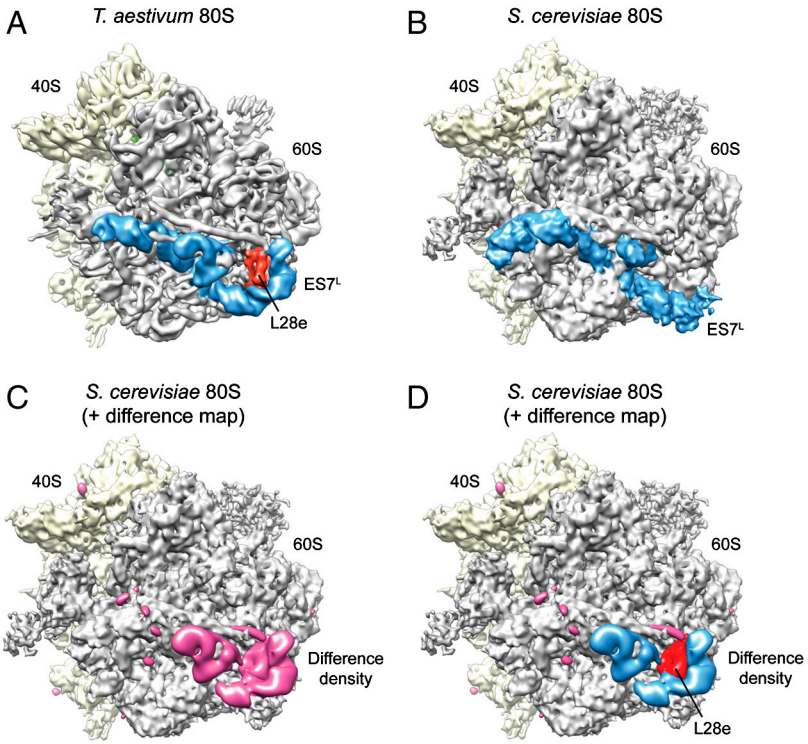


Fig. S2. Difference maps generated between *Triticum aestivum* and *Saccharomyces cerevisiae* identify position of r-protein L28e. (*A*) *T. aestivum* 80S reconstruction with ES7^{L} (blue) and L28e (red) positions highlighted. (*B*) *S. cerevisiae* 80S reconstruction with ES7^{L} (blue) highlighted. (*C*) *S. cerevisiae* 80S reconstruction superimposed with the difference density (magenta) calculated between the (*A*) *T. aestivum* map and (*B*) the *S. cerevisiae* map. (*D*) Same as *C* but with the regions of the difference density corresponding to ES7^{L} in *T. aestivum* colored blue, leaving a large region of extra density (red) that was assigned as r-protein L28e.

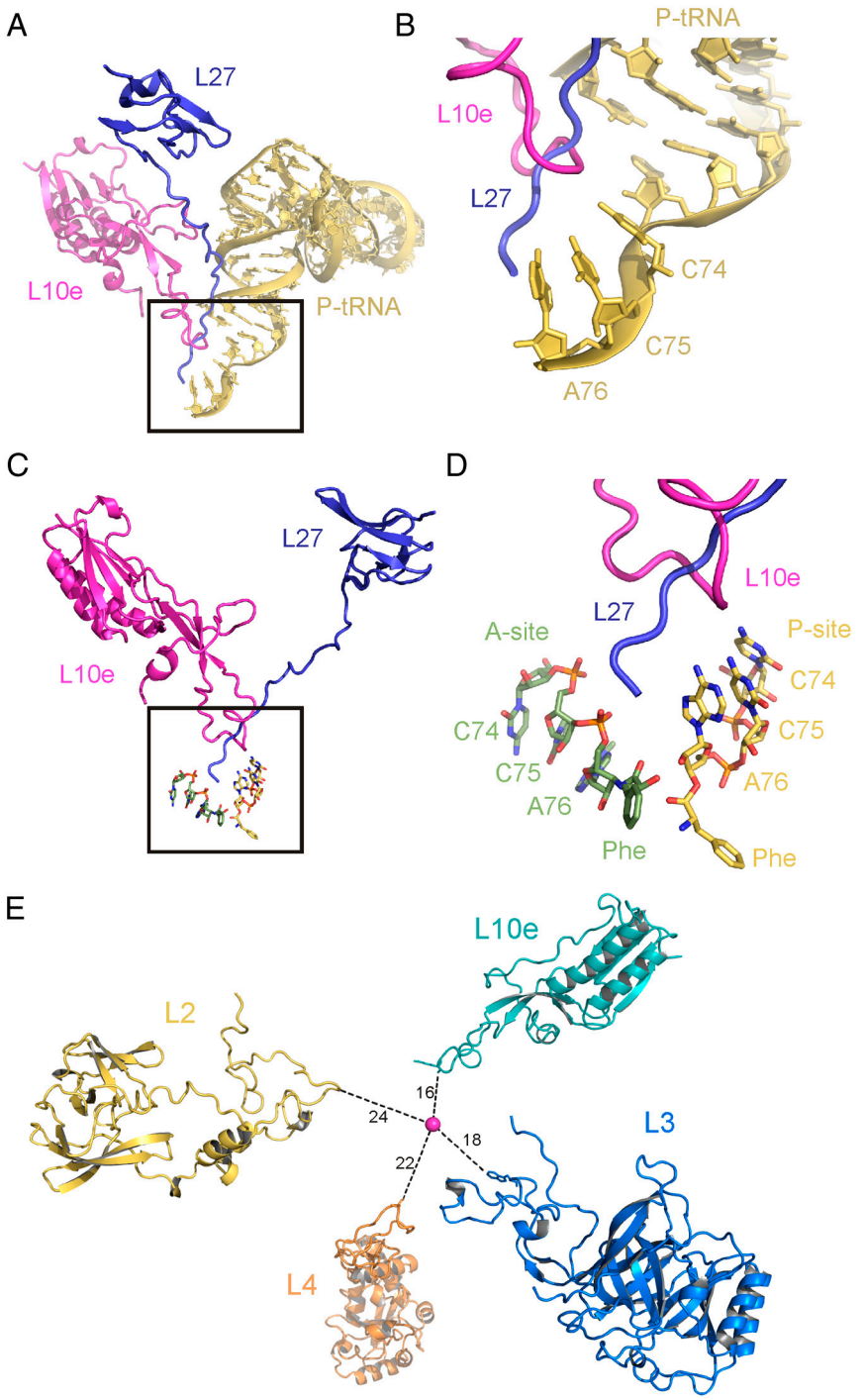


Fig. S3. Ribosomal proteins that approach the peptidyltransferase center of the ribosome. (*A* and *B*) Comparison of the relative positions of the N terminus of bacterial r-protein L27 (blue) and eukaryotic L10e (magenta) with a tRNA in the P site (yellow). (*C* and *D*) Comparison of the relative positions of the N terminus of bacterial r-protein L27 (blue) (1) and eukaryotic L10e (magenta) with the CCA-ends of tRNA mimics in A- (green) and P site (yellow) (2). (*E*) R-proteins L2p (yellow), L3p (blue), L4p (orange), and L10e (aqua) come within approximately 24, 22, 18, and 16 Å of the site of peptide bond formation, based on ref. 3.

1 Voorhees RM, et al. (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* 16:528–533.
2 Hansen JL, Schmeing TM, Moore PB, Steitz TA (2002) Structural insights into peptide bond formation. *Proc Natl Acad Sci USA* 99:11670–11675.
3 Nissen P, et al. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289:920–930.

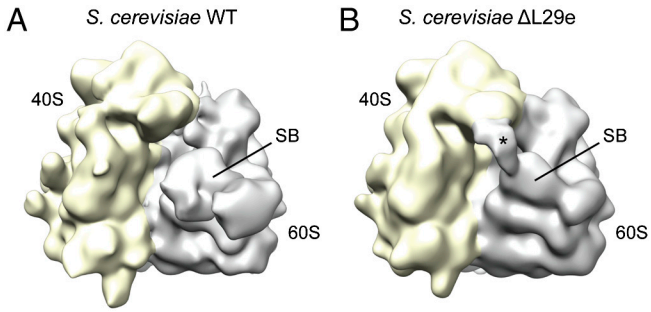


Fig. S4. Localization of ribosomal protein L29e. Reconstruction of (*A*) *Saccharomyces cerevisiae* WT 80S ribosome, compared to (*B*) reconstruction of *S. cerevisiae* 80S ribosomes isolated from a strain lacking the gene for L29e. In *B*, the rearranged position of the stalk base (SB) on the large subunit (gray) leads to a contact between the stalk (*) and the head of the small 40S subunit (yellow).

Table S1. Nomenclature for r proteins of the *Saccharomyces cerevisiae* and *Triticum aestivum*

	Family name	*S. cerevisiae name*	*Oryza sativa name*
1	S2p	rpS0	Sa
2	S3p	rpS3	S3
3	S3ae	rpS1	S3a
4	S4p	rpS9	S9
5	S4e	rpS4	S4
6	S5p	rpS2	S2
7	S6e	rpS6	S6
8	S7p	rpS5	S5
9	S7e	rpS7	S7
10	S8p	rpS22	S15a
11	S8e	rpS8	S8
12	S9p	rpS16	S16
13	S10p	rpS20	S20
14	S10e	rpS10	S10
15	S11p	rpS14	S14
16	S12p	rpS23	S23
17	S12e	rpS12	S12
18	S13p	rpS18	S18
19	S14p	rpS29	S29
20	S15p	rpS13	S13
21	S17p	rpS11	S11
22	S17e	rpS17	S17
23	S19p	rpS15	S15
24	S19e	rpS19	S19
25	S21e	rpS21	S21
26	S24e	rpS24	S24
27	S25e	rpS25	S25
28	S26e	rpS26	S26
29	S27e	rpS27	S27
30	S27ae	rpS31	S27a
31	S28e	rpS28	S28
32	S30e	rpS30	S30
33	RACK1	RACK1	RACK1

	Family name	*S. cerevisiae name*	*Oryza sativa name*
	Family name	*S. cerevisiae name*	*Triticum aestivum name*
1	L1p	rpL1	L1
2	L2p	rpL2	L2
3	L3p	rpL3	L3
4	L4e/L4p	rpL4	L4
5	L5p	rpL11	L11
6	L6p	rpL9	L9
7	L6e	rpL6	L6
8	L7ae	rpL8	L7a
9	L10p	rpP0	P0
10	L10e	rpL10	L10
11	L11p	rpL12	L12
12	L12p	rpP1/rpP2	P1/P2
13	L13p	rpL16	L13a
14	L13e	rpL13	L13
15	L14p	rpL23	L23
16	L14e	rpL14	L14
17	L15p	rpL28	L27a
18	L15e	rpL15	L15
19	L18p	rpL5	L5
20	L18e	rpL18	L18
21	L18ae	rpL20	L18a
22	L19e	rpL19	L19
23	L21e	rpL21	L21
24	L22p	rpL17	L17
25	L22e	rpL22	L22
26	L23p	rpL25	L23a
27	L24p	rpL26	L26
28	L24e	rpL24	L24
29	L27e	rpL27	L27
30	L28e	—	L28
31	L29p	rpL35	L35
32	L29e	rpL29	L29
33	L30p	rpL7	L7
34	L30e	rpL30	L30
35	L31e	rpL31	L31
36	L32e	rpL32	L32
37	L34e	rpL34	L34
38	L35ae	rpL33	L35a
39	L36e	rpL36	L36
40	L37e	rpL37	L37
41	L37ae	rpL43	L37a
42	L38e	rpL38	L38
43	L39e	rpL39	L39
44	L40e	rpL40	L40
45	L41e	rpL41	L41
46	L44e	rpL42	L44

Table S2. Summary of modeled wheat germ small subunit r proteins

Protein name	Protein family	Organism	Acc. no.	Size, aa	Modeled length, aa	Modeled range, aa	Percent modeled, %	Template	PDB ID
Sa	S2p	*Oryza sativa*	Q8H3I3	305	260	1–260	85	*Thermus thermophilus*	2J00_B
S2	S5p	*Oryza sativa*	Q84M35	274	263	1–263	96	*Escherichia coli*	
2QAL_E	S3	S3p	*Triticum aestivum*		Q8L804	227	208	12–219	92
Escherichia coli	2QAL_C								
S4*	**S4e**	***Oryza sativa***	**P49398**	**265**	**200**	**43–242**	**75**	***Thermoplasma acidophilum***	
3KBG_A	S5	S7p	*Oryza sativa*		Q93VC6	200	191	10–200	96
Pyrococcus horikoshii	1IQV_A								
S7	**S7e**	***Triticum aestivum***	**Q5I7K2**	**192**	**143**	**1–143**	**74**	***Polyalanine***	—
S9	S4p	*Oryza sativa*	Q2R1J8	195	195	1–195	100	*Thermus thermophilus*	2J00_D
S11	S17p	*Oryza sativa*	Q7XIK5	161	85	40–124	53	*Thermus thermophilus*	2J00_Q
S13	S15p	*Oryza sativa*	Q69UI2	151	121	31–151	80	*Escherichia coli*	
2QAL_O	S14	S11p	*Oryza sativa*		Q6H7T1	150	119	32–150	79
Thermus thermophilus	2J00_K								
S15	S19p	*Oryza sativa*	P31674	154	91	58–148	59	*Escherichia coli*	
2QAL_S	S16	S9p	*Oryza sativa*		Q0IQF7	149	126	24–149	85
Thermus thermophilus	2J00_I								
S17	**S17e**	***Oryza sativa***	**Q7XEQ3**	**141**	**141**	**1–141**	**100**	***Methanobacterium thermoautotrophicum***	
1RQ6_A	S18	S13p	*Triticum aestivum*		Q8L806	152	152	1–152	100
Escherichia coli	2QAL_M								
S19	**S19e**	***Oryza sativa***	**P40978**	**146**	**146**	**1–146**	**100**	***Pyrococcus abyssi***	
2V7F_A	S20	S10p	*Oryza sativa*	P35686	128	128	1–128	100	
Thermus thermophilus	2J00_J								
S21	**S21e**	***Oryza sativa***	**P35687**	**82**	**82**	**1–82**	**100**	**de novo**	—
S15a	S8p	*Oryza sativa*	Q84AP1	130	130	1–130	100	*Escherichia coli*	
2QAL_H	S23	S12p	*Oryza sativa*		Q8L4F2	142	142	1–142	100%
Thermus thermophilus	2J00_L								
S24	**S24e**	***Oryza sativa***	**Q6H541**	**138**	**98**	**5–102**	**71**	***Pyrococcus abyssi***	
2V94_A	**S25**	**S25e**	***Oryza sativa***		**Q53QG2**	**108**	**100**	**9–108**	**93**
Pyrococcus horikoshii	**1UB9_A**								
S26	**S26e**	***Oryza sativa***	**P49216**	**133**	**92**	**1–31; 73–133**	**69**	***Polyalanine***	—
S27	**S27e**	***Oryza sativa***	**Q6K5R5**	**86**	**50**	**33–82**	**58**	***Archeoglobus fulgidus***	
1QXF_A	**S28**	**S28e**	***Triticum aestivum***		**Q7X9K4**	**65**	**58**	**1–58**	**89**
Pyrococcus horikoshii	**1NY4_A**								
S29	S14p	*Triticum aestivum*	Q5I7K3	56	48	9–56	86	*Thermus thermophilus*	2J00_N
S30	**S30e**	***Oryza sativa***	**Q6K853**	**62**	**62**	**1–62**	**100**	**de novo**	—
RACK1	RACK1	*Triticum aestivum*	Q8LNY6	380	380	1–380	100	*Mus musculus*	2PBI_B

Acc., accession; PDB, Protein Data Bank.

*Bold rows indicate newly localized r proteins in this study, compared to the bacterial and archaeal X-ray structures.

Table S3. Summary of modeled wheat germ large subunit r proteins

Protein name	Protein family	Organism	Acc. no.	Size, aa	Modeled length	Modeled range	Percent modeled, %	Template	PDB ID
L1	L1p	*Triticum aestivum*	Q5I7L3	216	216	1–216	100	*Thermus thermophilus*	2HW8_A
L2	L2p	*Oryza sativa*	Q2QNF3	261	255	1–255	98	*Haloarcula marismortui*	1VQ8_A
L3	L3p	*Triticum aestivum*	Q7X744	389	389	1–389	100	*Haloarcula marismortui*	1VQ8_B
L4	L4p/L4e	*Oryza sativa*	Q6ZLB8	405	372	1–269;303–405	92	*Haloarcula marismortui*	1VQ8_C
L5	L18p	*Oryza sativa*	Q8L4L4	304	304	1–304	100	*Haloarcula marismortui*	1VQ8_N
L6*	**L6e**	***Triticum aestivum***	**Q5I7L4**	**219**	**219**	**1–219**	**100**	***Sulfolobus solfataricus***	**2JOY_A**
L7	L30p	*Triticum aestivum*	Q5I7K6	244	244	1–244	100	*Haloarcula marismortui*	1VQ8_W
L7a	L7ae	*Oryza sativa*	P35685	258	201	58–258	78	*Haloarcula marismortui*	1VQ8_F
L9	L6p	*Oryza sativa*	P49210	190	190	1–190	100	*Haloarcula marismortui*	1VQ8_E
L10	L10e	*Oryza sativa*	Q0ITS8	224	192	33–224	86	*Haloarcula marismortui*	3CC2_H
L11	L5p	*Triticum aestivum*	Q5I7L2	180	170	1–170	94	*H. marismortui/T. thermophilus*	1VQ8_D/2J01_G
L12	L11p	*Oryza sativa*	Q0JAI2	166	128	12–139	77	*Haloarcula marismortui*	2QA4_I
L13	**L13e**	***Oryza sativa***	**Q7XJB4**	**208**	**182**	**13–194**	**88**	***Polyalanine***	**—**
L14	**L14e**	***Oryza sativa***	**Q7XJ52**	**134**	**134**	**1–134**	**100**	***Sulfolobus solfataricus***	**2JOY_A**
L15	L15e	*Oryza sativa*	Q8H8S1	204	194	1–194	95	*Haloarcula marismortui*	3CC2_M
L13a	L13p	*Triticum aestivum*	Q5I7L1	206	206	1–206	100	*Haloarcula marismortui*	1VQ8_J
L17	L22p	*Oryza sativa*	Q6ZIA1	171	171	1–171	100	*Haloarcula marismortui*	1VQ8_R
L18	L18e	*Triticum aestivum*	Q5I7L0	188	163	1–163	87	*Haloarcula marismortui*	1VQ8_O
L18a	**L18ae**	***Oryza sativa***	**Q7XY20**	**178**	**167**	**1–167**	**94**	***Methanobacterium thermoautothropicum***	**2JXT_A**
L19	L19e	*Triticum aestivum*	Q943F3	209	189	1–189	90	*Haloarcula marismortui*	1VQ8_P
L21	L21e	*Triticum aestivum*	Q7XYC9	164	164	1–164	100	*Haloarcula marismortui*	1VQ8_Q
L22	**L22e**	***Oryza sativa***	**Q6YSX0**	**130**	**108**	**14–121**	**83**	***Artificial gene***	**2KL8_A**
L23	L14p	*Triticum aestivum*	Q5I7K4	140	140	1–140	100	*Haloarcula marismortui*	1VQ8_K
L24	L24e	*Oryza sativa*	Q5N754	162	75	1–75	46	*Haloarcula marismortui*	1VQ8_U
L23a	L23p	*Oryza sativa*	Q0JBZ7	152	122	31–152	80	*Haloarcula marismortui*	1VQ8_S
L26	L24p	*Oryza sativa*	Q2QXN5	150	130	1–130	87	*Haloarcula marismortui*	1VQ8_T
L27a	L15p	*Oryza sativa*	Q6EUQ7	144	144	1–144	100	*Haloarcula marismortui*	1VQ8_L
L27	**L27e**	***Oryza sativa***	**Q7XC31**	**136**	**99**	**1–99**	**73**	***Sulfolobus solfataricus***	**2JOY_A**
L28	**L28e**	***Oryza sativa***	**Q5TKP3**	**147**	**73**	**58–130**	**50**	**de novo**	**—**
L29	**L29e**	***Oryza sativa***	**Q9FP55**	**60**	**23**	**38–60**	**38**	***Oryctolagus cuniculus***	**1UTG_A**
L30	L30e	*Triticum aestivum*	Q5I7K9	112	112	1–112	100	*Saccharomyces cerevisiae*	1CN7_A
L31	L31e	*Triticum aestivum*	Q6ZGV5	123	120	1–120	98	*Haloarcula marismortui*	1VQ8_X
L32	L32e	*Oryza sativa*	Q3MST7	133	133	1–133	100	*Haloarcula marismortui*	1VQ8_Y
L34	**L34e**	***Triticum aestivum***	**Q5I7K8**	**119**	**119**	**1–119**	**100**	***Rhodobacter capsulatus***	**2PPT_A**
L35a	**L35ae**	***Oryza sativa***	**Q6I608**	**111**	**104**	**1–104**	**94**	***Pyrococcus furiosus***	**1SQR_A**
L35	L29p	*Triticum aestivum*	Q8L805	124	124	1–124	100	*Haloarcula marismortui*	1VQ8_V
L36	**L36e**	***Triticum aestivum***	**Q5I7L5**	**112**	**77**	**27–103**	**69**	***Archeoglobus fulgidus***	**2OEB_A**
L37	L37e	*Oryza sativa*	Q6Z8Y5	94	94	1–94	100	*Haloarcula marismortui*	1VQ8_1
L38	**L38e**	***Oryza sativa***	**Q8GVY2**	**69**	**69**	**1–69**	**100**	***Homo sapiens***	**1WH9_A**
L39	L39e	*Triticum aestivum*	Q5I7K7	51	51	1–51	100	*Haloarcula marismortui*	1VQ8_2
L40	**L40e**	***Oryza sativa***	**P35296**	**53**	**41**	**13–53**	**77**	***Sulfolobus solfataricus***	**2AYJ_A**
L41	**L41e**	***Oryza sativa***	**P62125**	**25**	**25**	**1–25**	**100**	**de novo**	**—**
L42	L44e	*Oryza sativa*	Q8H5N0	105	105	1–105	100	*Haloarcula marismortui*	1VQ8_3
L43	L37ae	*Oryza sativa*	Q5QM99	92	92	1–92	100	*Haloarcula marismortui*	3CC2_Z
P0	L10p	*Oryza sativa*	P41095	319	262	1–262	82	*Pyrococcus horikoshii and Methanocaldococcus janaschii*	3A1Y_G and 3JSY_A
P1	L12p	*Triticum aestivum*	Q5I7K5	110	58	6–63	53	*Pyrococcus horikoshii*	3A1Y_E
P2	L12p	*Triticum aestivum*	Q7X729	112	59	1–59	53	*Pyrococcus horikoshii*	3A1Y_F

Acc., accession; PDB, Protein Data Bank.

*Bold rows indicate newly localized r proteins in this study, compared to the bacterial and archaeal X-ray structures.

Table S4. Summary of modeled yeast small subunit r proteins

Protein name	Protein family	Organism	Acc. no.	Size, aa	Modeled length, aa	Modeled range, aa	Percent modeled, %	Template	PDB ID
rpS0	S2p	*Saccharomyces cerevisiae*	P32905	252	252	1–252	100	*Thermus thermophilus*	2J00_B
rpS2	S5p	*Saccharomyces cerevisiae*	P25443	254	254	1–254	100	*Escherichia coli*	2QAL_E
rpS3	S3p	*Saccharomyces cerevisiae*	P05750	240	204	12–215	85	*Escherichia coli*	2QAL_C
rpS4*	**S4e**	***Saccharomyces cerevisiae***	**P05753**	**261**	**200**	**43–242**	**77**	***Thermoplasma acidophilum***	**3KBG_A**
rpS5	S7p	*Saccharomyces cerevisiae*	P26783	225	199	27–225	88	*Pyrococcus horikoshii*	1IQV_A
rpS7	**S7e**	***Saccharomyces cerevisiae***	**P26786**	**190**	**143**	**1–143**	**75**	***Polyalanine***	**—**
rpS9	S4p	*Saccharomyces cerevisiae*	O13516	197	197	1–197	100	*Thermus thermophilus*	2J00_D
rpS11	S17p	*Saccharomyces cerevisiae*	P26781	156	85	39–123	54	*Thermus thermophilus*	2J00_Q
rpS13	S15p	*Saccharomyces cerevisiae*	P05756	151	121	31–151	80	*Escherichia coli*	2QAL_O
rpS14	S11p	*Saccharomyces cerevisiae*	P06367	137	119	19–137	87	*Thermus thermophilus*	2J00_K
rpS15	S19p	*Saccharomyces cerevisiae*	Q01855	142	88	49–136	62	*Escherichia coli*	2QAL_S
rpS16	S9p	*Saccharomyces cerevisiae*	P40213	143	126	18–143	88	*Thermus thermophilus*	2J00_I
rpS17	**S17e**	***Saccharomyces cerevisiae***	**P02407**	**136**	**136**	**1–136**	**100**	***Methanobacterium thermoautotrophicum***	**1RQ6_A**
rpS18	S13p	*Saccharomyces cerevisiae*	P35271	146	140	7–146	96	*Escherichia coli*	2QAL_M
rpS19	**S19e**	***Saccharomyces cerevisiae***	**P07280**	**144**	**144**	**1–144**	**100**	***Pyrococcus abyssi***	**2V7F_A**
rpS20	S10p	*Saccharomyces cerevisiae*	P38701	121	113	9–121	93	*Thermus thermophilus*	2J00_J
rpS21	**S21e**	***Saccharomyces cerevisiae***	**P0C0V8**	**87**	**87**	**1–87**	**100**	**de novo**	**—**
rpS22	S8p	*Saccharomyces cerevisiae*	P0C0W1	130	130	1–130	100	*Escherichia coli*	2QAL_H
rpS23	S12p	*Saccharomyces cerevisiae*	P32827	145	145	1–145	100	*Thermus thermophilus*	2J00_L
rpS24	**S24e**	***Saccharomyces cerevisiae***	**P26782**	**135**	**96**	**1–96**	**71**	***Pyrococcus abyssi***	**2V94_A**
rpS25	**S25e**	***Saccharomyces cerevisiae***	**Q3E792**	**108**	**85**	**24–108**	**78**	***Pyrococcus horikoshii***	**1UB9_A**
rpS26	**S26e**	***Saccharomyces cerevisiae***	**P39938**	**119**	**92**	**1–31;59–119**	**77**	**de novo**	**—**
rpS27	**S27e**	***Saccharomyces cerevisiae***	**P35997**	**82**	**50**	**31–80**	**61**	***Archeoglobus fulgidus***	**1QXF_A**
rpS28	**S28e**	***Saccharomyces cerevisiae***	**Q3E7X9**	**67**	**60**	**1–60**	**90**	***Pyrococcus horikoshii***	**1NY4_A**
rpS29	S14p	*Saccharomyces cerevisiae*	P41057	56	48	9–56	86	*Thermus thermophilus*	2J00_N
rpS30	**S30e**	***Saccharomyces cerevisiae***	**Q12087**	**63**	**63**	**1–63**	**100**	**de novo**	**—**
RACK1	RACK1	*Saccharomyces cerevisiae*	P38011	319	319	1–319	100	*Mus musculus*	2PBI_B

Acc., accession; PDB, Protein Data Bank.

*Bold rows indicate newly localized r proteins in this study, compared to the bacterial and archaeal X-ray structures.

Table S5. Summary of modeled yeast large subunit r proteins

Protein name	Protein family	Organism	Acc.no.	Size, aa	Modeled length	Modeled range	Percent modeled, %	Template	PDB ID
rpL1	L1p	*Saccharomyces cerevisiae*	P53030	217	217	1–217	100	*Thermus thermophilus*	2HW8_A
rpL2	L2p	*Saccharomyces cerevisiae*	P05736	254	254	1–254	100	*Haloarcula marismortui*	1VQ8_A
rpL3	L3p	*Saccharomyces cerevisiae*	P14126	387	387	1–387	100	*Haloarcula marismortui*	1VQ8_B
rpL4	L4p/L4e	*Saccharomyces cerevisiae*	P10664	362	329	1–261 ; 295–362	91	*Haloarcula marismortui*	1VQ8_C
rpL5	L18p	*Saccharomyces cerevisiae*	P26321	297	297	1–297	100	*Haloarcula marismortui*	1VQ8_N
rpL6*	**L6e**	***Saccharomyces cerevisiae***	**Q02326**	**176**	**176**	**1–176**	**100**	***Sulfolobus solfataricus***	**2JOY_A**
rpL7	L30p	*Saccharomyces cerevisiae*	P05737	244	239	6–244	98	*Haloarcula marismortui*	1VQ8_W
rpL8	L7ae	*Saccharomyces cerevisiae*	P17076	256	197	60–256	77	*Haloarcula marismortui*	1VQ8_F
rpL9	L6p	*Saccharomyces cerevisiae*	P05738	191	191	1–191	100	*Haloarcula marismortui*	1VQ8_E
rpL10	L10e	*Saccharomyces cerevisiae*	P41805	221	189	33–221	86	*Haloarcula marismortui*	3CC2_H
rpL11	L5p	*Saccharomyces cerevisiae*	P0C0W9	174	168	1–168	96	*Haloarcula marismortui and Thermus thermophilus*	1VQ8_D and 2J01_G
rpL12	L11p	*Saccharomyces cerevisiae*	P17079	165	127	12–138	77	*Haloarcula marismortui*	2QA4_I
rpL13	**L13e**	***Saccharomyces cerevisiae***	**Q12690**	**199**	**169**	**14–182**	**85**	***Polyalanine***	**—**
rpL14	**L14e**	***Saccharomyces cerevisiae***	**P36105**	**138**	**138**	**1–138**	**100**	***Sulfolobus solfataricus***	**2JOY_A**
rpL15	L15e	*Saccharomyces cerevisiae*	P05748	204	193	1–193	95	*Haloarcula marismortui*	3CC2_M
rpL16	L13p	*Saccharomyces cerevisiae*	P26784	199	199	1–199	100	*Haloarcula marismortui*	1VQ8_J
rpL17	L22p	*Saccharomyces cerevisiae*	P05740	184	170	1–170	92	*Haloarcula marismortui*	1VQ8_R
rpL18	L18e	*Saccharomyces cerevisiae*	P07279	186	161	1–161	87	*Haloarcula marismortui*	1VQ8_O
rpL19	L19e	*Saccharomyces cerevisiae*	P05735	189	189	1–189	100	*Haloarcula marismortui*	1VQ8_P
rpL20	**L18ae**	***Saccharomyces cerevisiae***	**P0C2I0**	**172**	**167**	**1–167**	**97**	***Methanobacterium thermoautothropicum***	**2JXT_A**
rpL21	L21e	*Saccharomyces cerevisiae*	Q02753	160	160	1–160	100	*Haloarcula marismortui*	1VQ8_Q
rpL22	**L22e**	***Saccharomyces cerevisiae***	**P05749**	**121**	**105**	**6–110**	**87**	***Artificial gene***	**2KL8_A**
rpL23	L14p	*Saccharomyces cerevisiae*	P04451	137	131	7–137	96	*Haloarcula marismortui*	1VQ8_K
rpL24	L24e	*Saccharomyces cerevisiae*	P04449	155	73	1–73	47	*Haloarcula marismortui*	1VQ8_U
rpL25	L23p	*Saccharomyces cerevisiae*	P04456	142	122	21–142	86	*Haloarcula marismortui*	1VQ8_S
rpL26	L24p	*Saccharomyces cerevisiae*	P05743	127	123	1–123	97	*Haloarcula marismortui*	1VQ8_T
rpL27	**L27e**	***Saccharomyces cerevisiae***	**P0C2H6**	**136**	**95**	**5–99**	**70**	***Sulfolobus solfataricus***	**2JOY_A**
rpL28	L15p	*Saccharomyces cerevisiae*	P02406	149	149	1–149	100	*Haloarcula marismortui*	1VQ8_L
rpL29	**L29e**	***Saccharomyces cerevisiae***	**P05747**	**59**	**22**	**38–59**	**37**	***Oryctolagus cuniculus***	**1UTG_A**
rpL30	L30e	*Saccharomyces cerevisiae*	P14120	105	105	1–105	100	*Saccharomyces cerevisiae*	1CN7_A
rpL31	L31e	*Saccharomyces cerevisiae*	P0C2H8	113	110	1–110	97	*Haloarcula marismortui*	1VQ8_X
rpL32	L32e	*Saccharomyces cerevisiae*	P38061	130	130	1–130	100	*Haloarcula marismortui*	1VQ8_Y
rpL33	**L35ae**	***Saccharomyces cerevisiae***	**P05744**	**107**	**100**	**1–100**	**93**	***Pyrococcus furiosus***	**1SQR_A**
rpL34	**L34e**	***Saccharomyces cerevisiae***	**P87262**	**121**	**118**	**1–118**	**97**	***Rhodobacter capsulatus***	**2PPT_A**
rpL35	L29p	*Saccharomyces cerevisiae*	P39741	120	118	3–120	98	*Haloarcula marismortui*	1VQ8_V
rpL36	**L36e**	***Saccharomyces cerevisiae***	**P05745**	**100**	**77**	**24–100**	**77**	***Archeoglobus fulgidus***	**2OEB_A**
rpL37	L37e	*Saccharomyces cerevisiae*	P49166	88	88	1–88	100	*Haloarcula marismortui*	1VQ8_1
rpL38	**L38e**	***Saccharomyces cerevisiae***	**P49167**	**78**	**78**	**1–78**	**100**	***Homo sapiens***	**1WH9_A**
rpL39	L39e	*Saccharomyces cerevisiae*	P04650	51	51	1–51	100	*Haloarcula marismortui*	1VQ8_2
rpL40	**L40e**	***Saccharomyces cerevisiae***	**P14796**	**52**	**40**	**13–52**	**77**	***Sulfolobus solfataricus***	**2AYJ_A**
rpL41	**L41e**	***Saccharomyces cerevisiae***	**P05746**	**25**	**25**	**1–25**	**100**	**de novo**	**—**
rpL42	L44e	*Saccharomyces cerevisiae*	P02405	106	106	1–106	100	*Haloarcula marismortui*	1VQ8_3
rpL43	L37ae	*Saccharomyces cerevisiae*	P49631	92	92	1–92	100	*Haloarcula marismortui*	3CC2_Z
rpP0	L10p	*Saccharomyces cerevisiae*	P05317	312	257	1–257	82	*Pyrococcus horikoshii and Methanocaldococcus janaschii*	3A1Y_G and 3JSY_A
rpP1	L12p	*Saccharomyces cerevisiae*	P05318	106	58	5–62	54	*Pyrococcus horikoshii*	3A1Y_E
rpP2	L12p	*Saccharomyces cerevisiae*	P05319	106	58	1–58	55	*Pyrococcus horikoshii*	3A1Y_F

*Bold rows indicate newly localized r proteins in this study, compared to the bacterial and archaeal X-ray structures.

Table S6. Localization of unassigned eukaryotic 80S r proteins

Protein family	*Triticum aestivum* name	*Saccharomyces cerevisiae* name	Localization basis
L6e	L6	rpL6	(*i*) Cross-linking and accessibility to proteolysis (1). (*ii*) Comparison of *S. cerevisiae* and *T. aestivum* cryo-EM maps (which contain L6e) with the *Haloarcula marismortui* 50S X-ray structure and the *Pyrococcus furiosus* 70S cryo-EM structure (which lack L6e). (*iii*) The N terminus of L6e was assigned based on differences between the density and the sequence of *T. aestivum* and *S. cerevisiae*.
L13e	L13	rpL13	(*i*) Cross-linking and accessibility to proteolysis (1). (*ii*) Heterogeneous distribution in archaea.
L14e	L14	rpL14	(*i*) Fold search and secondary structure prediction. (*ii*) Length differences between *S. cerevisiae* and *T. aestivum* L14e sequences, i.e., C terminus is longer in *T. aestivum* L14e and N terminus is longer in *S. cerevisiae*.
L18ae	L18a	rpL20	(*i*) The difference between archaea and eukaryotes, namely, that majority of this density existed only on eukaryotic ribosomes. (*ii*) Fold search revealed that the protein consists of two domains with a distinct LX motif.
L22e	L22	rpL22	(*i*) Cross-linking and accessibility to proteolysis (1). (*ii*) Comparison of *S. cerevisiae* and *T. aestivum* cryo-EM maps (which contain L22e) with the *H. marismortui* 50S X-ray structure and the *P. furiosus* 70S cryo-EM structure (which lack L22e).
L27e	L27	rpL27	(*i*) Cross-linking and accessibility to proteolysis (1). (*ii*) Comparison of *S. cerevisiae* and *T. aestivum* cryo-EM maps (which contain L27e) with the *H. marismortui* 50S X-ray structure and the *P. furiosus* 70S cryo-EM structure (which lack L27e).
L28e	L28	—	(*i*) Difference map between *S. cerevisiae* 80S cryo-EM map with *T. aestivum* 80S cryo-EM maps, because L28e does not exist in *S. cerevisiae*, but is present in *T. aestivum* (2). (*ii*) Cross-linking and accessibility to proteolysis (1).
L29e	L29	rpL29	(*i*) Cryo-EM reconstruction of ΔL29e-80S ribosome at 20.5 Å and comparison with yeast 80S ribosome from wild-type strain (Fig. 1 *F–H*).
L34e	L34	rpL34	Based on the fact that it exists in Eukarya and Archaea, on the fold of the model, and the fact that this was the only major density left unassigned.
L35ae	L35a	rpL33	(*i*) Cross-linking and accessibility to proteolysis (1). (*ii*) Heterogeneous distribution in Archaea.
L36e	L36	rpL36	(*i*) Cross-linking and accessibility to proteolysis (1). (*ii*) Comparison of *S. cerevisiae* and *T. aestivum* cryo-EM maps (which contain L36e) with the *H. marismortui* 50S X-ray structure and the *P. furiosus* 70S cryo-EM structure (which lack L36e).
L38e	L38	rpL38	(*i*) Cryo-EM reconstruction of ΔL38e-80S ribosome at 21 Å and comparison with yeast 80S ribosome from wildtype strain (Fig. 1 *C–E*).
L40e	L40	rpL40	(*i*) Fold and the size of the protein.
L41e	L41	rpL41	(*i*) Size and density features: L41e is only 25 amino acids. (*ii*) Location: isolated density that is unlikely to be an RNA or an r-protein extension.
S4e	S4	rpS4	(*i*) Cross-linking (3). (*ii*) Structural information from *Thermoplasma acidophilum* PDB 3KBG.
S7e	S7	rpS7	(*i*) Immuno-EM (4).
S17e	S17	rpS17	(*i*) Cross-linking (3). (*ii*) Localization: All remaining density on the head of the small subunit was already assigned. (*iii*) Structural information (5).
S19e	S19	rpS19	(*i*) Location: Assembly precursors indicate S19e to be associated with the head of the small subunit (6). (*ii*) Structural information (7). (*iii*) Subsequently localized in the fungi 80S ribosome (8).

Protein family	*Triticum aestivum* name	*Saccharomyces cerevisiae* name	Localization basis
S21e	S21	rpS21	(*i*) Immuno-EM (4).
S24e	S24	rpS24	(*i*) Immuno-EM (4).
S25e	S25	rpS25	(*i*) Cross-linking to IRES elements (9).
S26e	S26	rpS26	(*i*) Cross-linking to mRNA (10).
S27e	S27	rpS27	(*i*) Structural information (11).
S28e	S28	rpS28	(*i*) Cross-linking to mRNA (10). (*ii*) Structural information (12).
S30e	S30	rpS30	(*i*) Cross-linking to mRNA (13, 14).

IRES, internal ribosome entry site.

1 Marion MJ, Marion C (1987) Localization of ribosomal proteins on the surface of mammalian 60S ribosomal subunits by means of immobilized enzymes. Correlation with chemical cross-linking data. *Biochem Biophys Res Commun* 149:1077–1083.
2 Lecompte O, et al. (2002) Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. *Nucleic Acids Res* 30:5382–5390.
3 Gross B, Westermann P, Bielka H (1983) Spatial arrangement of proteins within the small subunit of rat liver ribosomes studied by cross-linking. *EMBO J* 2:255–260.
4 Bommer UA, Lutsch G, Stahl J, Bielka H (1991) Eukaryotic initiation factors eIF-2 and eIF-3: Interactions, structure and localization in ribosomal initiation complexes. *Biochimie* 73:1007–1019.
5 Wu B, et al. (2008) The solution structure of ribosomal protein S17E from *Methanobacterium thermoautotrophicum*: A structural homolog of the FF domain. *Protein Sci* 17:583–588.
6 Ferreira-Cerca S, et al. (2007) Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins. *Mol Cell* 28:446–457.
7 Gregory LA, et al. (2007) Molecular basis of Diamond-Blackfan anemia: Structure and function analysis of RPS19. *Nucleic Acids Res* 35:5913–5921.
8 Taylor DJ, et al. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure* 17:1591–1604.
9 Nishiyama T, Yamamoto H, Uchiumi T, Nakashima N (2007) Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. *Nucleic Acids Res* 35:1514–1521.
10 Pisarev AV, et al. (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* 27:1609–1621.
11 Herve du Penhoat C, et al. (2004) The NMR solution structure of the 30S ribosomal protein S27e encoded in gene RS27_ARCFU of *Archaeoglobus fulgidis* reveals a novel protein fold. *Protein Sci* 13:1407–1416.
12 Aramini JM, et al. (2003) Solution NMR structure of the 30S ribosomal protein S28E from *Pyrococcus horikoshii*. *Protein Sci* 12:2823–2830.
13 Bulygin K, et al. (2005) The first position of a codon placed in the A site of the human 80S ribosome contacts nucleotide C1696 of the 18S rRNA as well as proteins S2, S3, S3a, S30, and S15. *Biochemistry* 44:2153–2162.
14 Takahashi Y, Mitsuma T, Hirayama S, Odani S (2002) Identification of the ribosomal proteins present in the vicinity of globin mRNA in the 40S initiation complex. *J Biochem* 132:705–711.

ARTICLE

doi:10.1038/nature12104

Structures of the human and *Drosophila* 80S ribosome

Andreas M. Anger[1]*, Jean-Paul Armache[1]*, Otto Berninghausen[1], Michael Habeck[2,3], Marion Subklewe[4], Daniel N. Wilson[1] & Roland Beckmann[1]

Protein synthesis in all cells is carried out by macromolecular machines called ribosomes. Although the structures of prokaryotic, yeast and protist ribosomes have been determined, the more complex molecular architecture of metazoan 80S ribosomes has so far remained elusive. Here we present structures of *Drosophila melanogaster* and *Homo sapiens* 80S ribosomes in complex with the translation factor eEF2, E-site transfer RNA and Stm1-like proteins, based on high-resolution cryo-electron-microscopy density maps. These structures not only illustrate the co-evolution of metazoan-specific ribosomal RNA with ribosomal proteins but also reveal the presence of two additional structural layers in metazoan ribosomes, a well-ordered inner layer covered by a flexible RNA outer layer. The human and *Drosophila* ribosome structures will provide the basis for more detailed structural, biochemical and genetic experiments.

Crystal structures of prokaryotic ribosomal particles have provided insights into protein biosynthesis at both a structural and a functional level[1]. In contrast to their bacterial counterparts, eukaryotic ribosomes are much larger and more complex; they contain approximately 2,650 nucleotides of additional rRNA in *H. sapiens* in the form of so-called expansion segments and 26 additional ribosomal proteins as well as 2,452 amino acids of ribosomal protein extensions[2–4]. Cryo-electron microscopy (cryo-EM)[5–7] and crystal structures[8–10] have elucidated the architecture of yeast, protist and plant ribosomes. In contrast, the limited resolution (9 to 20 Å) of cryo-EM structures of mammalian 80S ribosomes[11–14] has so far prohibited the generation of complete molecular models for these metazoans.

Here we present single-particle cryo-EM structures of monomeric 80S ribosomes isolated from *D. melanogaster* embryonic extracts and human peripheral blood mononuclear cells (Supplementary Fig. 1). *In silico* sorting was used to generate homogeneous data sets with additional density corresponding to eukaryotic elongation factor 2 (eEF2), in agreement with mass spectrometry analysis (Supplementary Tables 1 and 2). The eEF2-containing particles seemed to be stabilized in a rotated conformation, allowing the reconstructions of each sub-data set to reach an average resolution of 5.4 to 6.0 Å (Supplementary Fig. 2). Notably, local resolution of the human 80S ribosome ranged from above 9 Å on the flexible periphery to better than 4.8 Å for large parts of the map (Fig. 1a). This is in agreement with the distinct structural details observed throughout the map: the pitch of α-helices is visible and strand-separation is recognizable for many β-sheets of ribosomal proteins (Fig. 1b). Density is also visible for a number of bulky side chains (Fig. 1b). In terms of rRNA, the phosphate–ribose backbone is well resolved and bulged-out bases are clearly represented (Fig. 1c). Moreover, the quality of the cryo-EM map enabled us to distinguish between human rRNA sequence variations (Supplementary Fig. 3). Our electron-density maps, coupled with secondary structure predictions for the rRNA expansion segments and the available yeast and *Tetrahymena* crystal structures[8–10], enabled us to build complete molecular models for both the *Drosophila* and human 80S ribosome (Fig. 1d, e and Supplementary Tables 3–8).

Ribosomal protein extensions

With the exception of yeast, which lacks L28e, eukaryotic cytoplasmic 80S ribosomes contain the same set of 80 core ribosomal proteins (Fig. 2a, b, Supplementary Fig. 4 and Supplementary Tables 3–6). Compared to yeast and protists, there has been a modest increase in protein mass in metazoan ribosomes, specifically by a total of 1,094 amino acids (approximately 8%) and 796 amino acids (approximately 6%) in the *Drosophila* and human 80S ribosomes, respectively. On the 40S subunit, the protein mass increase of *Drosophila* (210 amino acids, approximately 4%) and human (147 amino acids, 3%) relative to yeast is small, and mostly disordered in the cryo-EM maps. Notable exceptions include the carboxy-terminal extension (CTE) of S26e, which reaches into the messenger RNA exit channel (Supplementary Fig. 5), and part of the CTE of S6e that bridges the right and left feet of the 40S subunit (Fig. 2b and Supplementary Fig. 5). Phosphorylation of the CTE of S6e by S6 kinase (S6K) is important for translation regulation, as well as glucose homeostasis and regulation of cell size in metazoans[15]. The S6K consensus recognition motif (RXRXXS), which was disordered in recent X-ray structures of the yeast 80S ribosome and *Tetrahymena* 40S subunit[8,10], adopts an α-helical conformation in the human 80S ribosome (Supplementary Fig. 5). The most dramatic increases in ribosomal protein extensions are seen on the 60S subunits for ribosomal proteins L4, L6e, L14e and L29e, as well as for *Drosophila* L22e and L23. Collectively, these account for 52% (460 amino acids) and 58% (375 amino acids) of the total protein mass gain in the *Drosophila* and human 60S subunit, respectively. Notably, the approximately 180- and 140-amino-acid extensions of L22e and L23, respectively, double the size of these *D. melanogaster* ribosomal proteins, compared to other non-insect species such as yeast and human (Supplementary Fig. 6). Structures of yeast and *Tetrahymena* ribosomes revealed a highly complex network of RNA–protein interactions between the eukaryote-specific ribosomal protein extensions and the rRNA expansion segments[2–4,7–10]. The dimensions of this RNA–protein layer has developed further in metazoan ribosomes, which is illustrated by the increasing size and complexity of the interaction between expansion segment 7L

[1]Gene Center, Department of Biochemistry and Center for integrated Protein Science Munich (CiPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. [2]Department of Empirical Inference, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany. [3]Department of Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstrasse 38, 72076 Tübingen, Germany. [4]Department of Internal Medicine III, Klinikum der Universität München and Clinical Cooperation Group Immunotherapy at the Helmholtz Institute Munich, Marchioninistrasse 15, 81377 Munich, Germany.
*These authors contributed equally to this work.

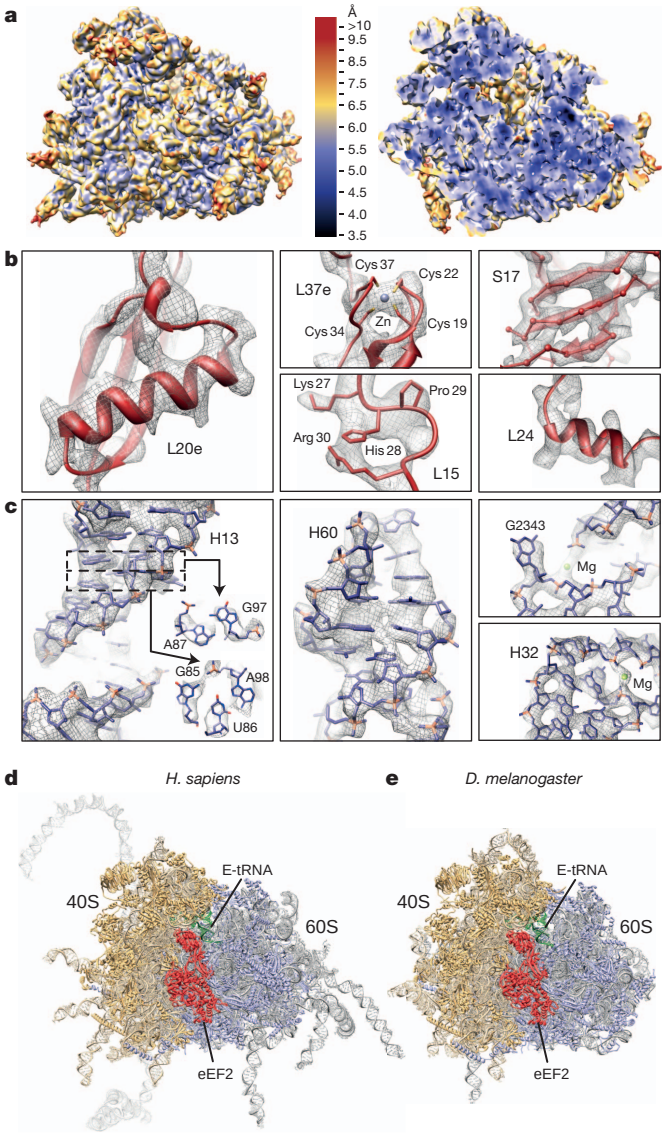


Figure 1 | Structures of the human and *Drosophila* 80S ribosomes. **a**, Surface and cross section of the human 80S ribosome electron density map (filtered at 6 Å for clarity), coloured according to the local resolution. **b**, **c**, Selected views of the *H. sapiens* 80S map (grey mesh) with (**b**) protein and (**c**) rRNA. RNA backbone phosphates are highlighted in orange. **d**, **e**, Complete models of the human and *Drosophila* 80S ribosomes with ribosomal proteins and rRNA of the 40S and 60S subunits shown in orange and blue, respectively. Flexible human ES27L (light grey) is shown in an arbitrary position.

(ES7L) with the NTE of L6e (Supplementary Fig. 7). Interestingly, the extensions of human L4, L14e and L29e and *Drosophila* L22e and L23, show similarity to the flexible C-terminal regions of the linker histone H1 in that they are highly basic and enriched in alanine, lysine and proline residues[16] (Supplementary Fig. 8). The histone H1 tails have been proposed to form α-helical conformations punctuated by proline breaks, which track one groove of the linker DNA (reviewed previously[17]). In the *Drosophila* and human 80S ribosome, it seems that these histone H1-like ribosomal protein parts are directed towards adjacent expansion segments. However, owing to the flexibility of the expansion segments, it was not possible to model the associated extensions (Supplementary Fig. 8).

Ribosomal proteins, eEF2 and Stm1-like factors

As in yeast and *Tetrahymena* extensions of *Drosophila* and human ribosomal proteins S13, S19, S25e, S30e and S31e extend into the functional centre of the 40S subunit. There, the amino-terminal extension (NTE) of ribosomal proteins S30e and S31e establish interactions with eEF2 (Supplementary Fig. 9). This was not observed in the lower resolution yeast eEF2–80S complexes[5,12,18]. Moreover, although the overall conformation and contacts of *Drosophila* and human eEF2 on the ribosome are very similar to those observed for yeast[5,12,18] (Fig. 2c, d, Supplementary Fig. 10, and Supplementary Tables 9 and 10), at higher resolution we could also model interactions between the N-terminal domain of L11, domain II of the L10 stalk protein and the G′ domain of human and *Drosophila* eEF2 (Fig. 2e). Additional density is present in the human eEF2 for the mammal-specific insertion within the G′ domain. This additional density is absent in *Drosophila* eEF2 (Fig. 2f). At lower thresholds, extra density is observed adjacent to this region. This may represent the C terminus of the 60S acidic ribosomal P1 and P2 stalk proteins, reminiscent of the interaction between the bacterial L7 and L12 stalk proteins and the G′ domain of EF-G[19,20]. In contrast to bacterial EF-G, archaeal EF2s and eEF2s are post-translationally modified by conversion of a conserved histidine (His 699, His 701 and His 715 for yeast, *Drosophila*, and human eEF2, respectively) to diphthamide. Deletion of the modification enzymes in mice leads to embryonic lethality or severe developmental defects[21]. Moreover, diphthamide is adenosine di-phosphate (ADP)-ribosylated by the diphtheria toxin, which inactivates eEF2 and inhibits protein synthesis[21]. In the human 80S–eEF2 structure, we observe density for the diphthamide residue contacting the backbone of H44 in the vicinity of A1825 (A1493 in *Escherichia coli* numbering) (Fig. 2g). In bacteria, A1492 and A1493 are involved in recognition of the mRNA–tRNA duplex during decoding[22,23], thus contact of diphthamide with this region is consistent with its proposed role to disrupt the interaction between the decoding centre and mRNA–tRNA duplex during translocation[18]. Notably, we also observe an alternative conformation of diphthamide directed towards density located within the path of the mRNA, which we have assigned to the serpine 1 mRNA-binding protein 1 (SERBP1; also known as plasminogen activator inhibitor 1 RNA-binding protein) based on mass-spectrometry analysis (Fig. 2c, g and Supplementary Table 1). SERBP1 was identified, together with ribosomal proteins and eIF3, to interact with the hepatitis C virus internal ribosomal entry site (IRES)[24], which engages the small ribosomal subunit during initiation[25]. Moreover, SERBP1 is homologous to the translation repressor Stm1 (ref. 26), which is present in the crystal structure of the yeast 80S ribosome purified under conditions of nutrient deprivation[10]. We observe that, like Stm1, SERBP1 has an extended structure passing through the P- and A-tRNA binding sites (Fig. 2c, d); it then follows the mRNA channel to the solvent side, where it interacts with ribosomal proteins S5, S10e, S12e and S30e located on the head of the 40S subunit (Fig. 2c and Supplementary Figs 10 and 11, and Supplementary Table 11). Examination of the *Drosophila* 80S ribosome also revealed a density within these regions, which was identified by mass spectrometry to be VIG2 (Supplementary Tables 2 and 12), a protein orthologous to SERBP1 (Supplementary Figs 10 and 11). The identification of SERBP1 and VIG2 on metazoan 80S ribosomes indicates a novel role, analogous to Stm1 in yeast, for these proteins in the regulation of translation in *Drosophila* and humans.

Ribosomal RNA expansion segments

We were able to localize and build models for all 30 rRNA expansion segments (we use an extended nomenclature based on a previous paper[27], Supplementary Fig. 12 and Supplementary Tables 7 and 8) of the *Drosophila* and human 80S ribosome, 9 expansion segments of the 40S subunit and 21 expansion segments of the 60S subunit (Fig. 3 and Supplementary Figs 13–16). Although human and *Drosophila* contain a similar set of expansion segments as yeast and protists, their expansion segments are generally much longer, exemplified by comparing ES3S, ES7L, ES9L, ES15L, ES27L and ES39L between yeast (approximately 110, 200, 70, 20, 160 and 140 nucleotides) and human (longer by 50, 670, 40, 170, 550 and 100 nucleotides) (Supplementary Tables 7 and 8). In addition, metazoans contain ES30L and ES43L (Fig. 3), which are lacking in yeast and *Tetrahymena*. Although the

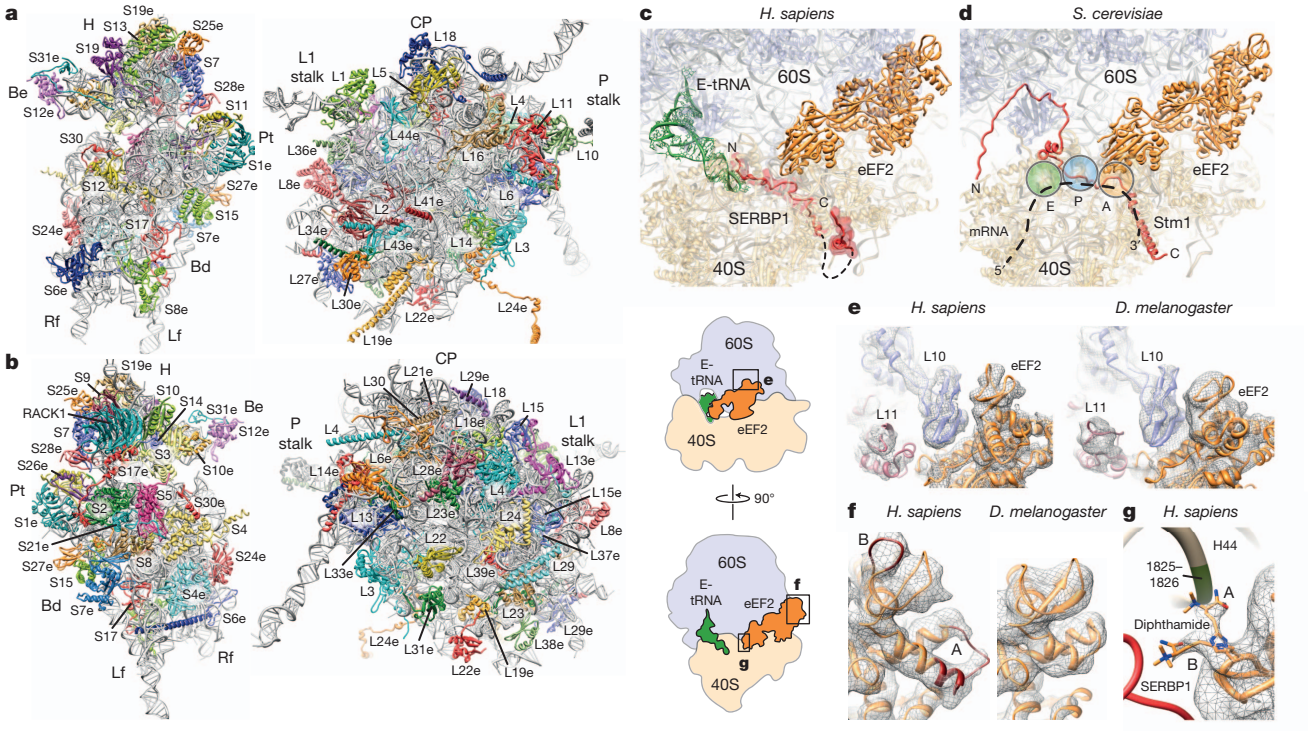


Figure 2 | Protein architecture of the human 80S ribosome and associated factors. **a**, **b**, Interface (**a**) and solvent (**b**) view of the human 40S (left) and 60S (right) ribosome subunits, with rRNA shown in grey and ribosomal proteins coloured individually. Be, beak; Bd, body; CP, central protuberance; H, head; Lf, left foot; Pt, platform; Rf, right foot. **c**, Relative position of eEF2 (orange), E-site tRNA (green) and SERBP1 (red) on the *H. sapiens* 80S ribosome **d**, eEF2[18] and Stm1 (red) in *S. cerevisiae*[10]. Positions of aminoacyl (A), peptidyl (P) and exit (E) tRNA-binding sites are indicated. **e**, Interaction of L11 and L10 with eEF2. **f**, G′ domains of eEF2 with human insertions (A and B, red). **g**, Alternative conformations of the diphthamide–His 715 of eEF2, contacting nucleotides 1825 and 1826 in H44 or SERBP1. The insets show the locations of the areas enlarged in parts **e**, **f** and **g**.

distal ends of several large human rRNA insertions (for example, ES3S, ES6S, ES7L, ES15L, ES27L, ES30L and ES39L) could only be partially resolved in the cryo-EM reconstructions (Supplementary Fig. 17), the flexible tentacle-like nature of these expansion segments was observable within individual electron-microscopy images (Supplementary Fig. 18)[28]. The extreme base composition of expansion segments, being AU-rich in *Drosophila* (32% GC) and GC-rich in human (80% GC) (Supplementary Tables 7, 8 and 13), has prevented secondary structure prediction for approximately 720 (33%) and 1,800 (57%) nucleotides of several expansion segments, respectively[29,30] (Supplementary Figs 19 and 20). However, using iterative model building and focused secondary structure predictions, we conclude that the distal ends of the flexible expansion segments adopt simple, unbranched A-form helices, enabling us to present complete molecular models

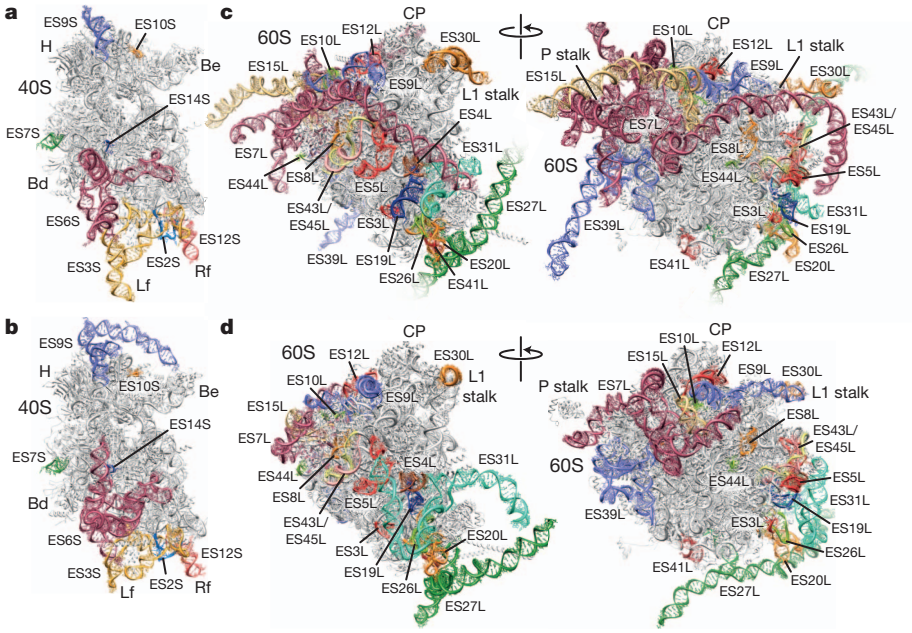


Figure 3 | Metazoan rRNA expansion segments. **a**, **b**, Molecular models of the 40S subunits of (**a**) *H. sapiens* and (**b**) *D. melanogaster* with expansion segments. **c**, **d**, Molecular models of the 60S ribosome subunits of (**c**) *H. sapiens* and (**d**) *D. melanogaster* showing expansion segments.

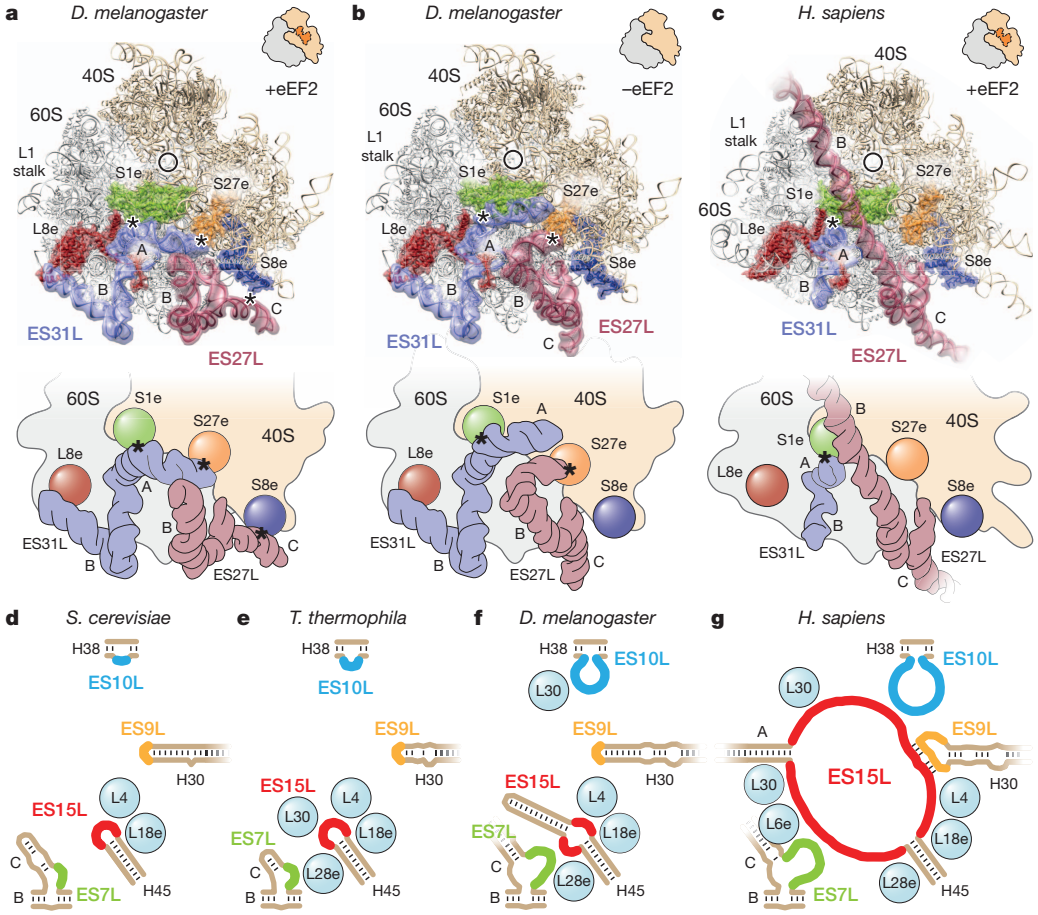


Figure 4 | Dynamic behaviour and co-evolution of expansion segments. **a–c**, Comparison of the ES27L and ES31L behaviour in the eEF2-bound (rotated) (**a**, **c**) and empty (-eEF2, unrotated) (**b**) form of the *Drosophila* and *H. sapiens* ribosome. Bridges with ribosomal proteins are highlighted with asterisks, the mRNA exit site is indicated with a circle. **d–g**, Schematic view comparing the interactions within the expansion-segment cluster formed by ES7L, ES9L, ES10L and ES15L between *S. cerevisiae* (**d**)[6,7,10], *T. thermophila* (**e**)[9], *D. melanogaster* (**f**) and *H. sapiens* (**g**). Non-helical elements of expansion segments are highlighted, and helices are labelled A to C.

(Fig. 3) and refined secondary structure diagrams for the entire human and *Drosophila* small and large subunit rRNAs (Supplementary Figs 13–16).

On the human and *Drosophila* 40S subunits, the expansion segments cluster at the bottom of the back of the particle, where ES3S and ES6S interact tightly (Fig. 3a, b). The terminal loop of helix E of ES6S (ES6S-E) forms continuous base pairs with an internal loop of ES3S-B (Supplementary Fig. 21), as reported previously for yeast, wheat germ and *Tetrahymena*[6,8,10,31]. ES3S-B is extended in human compared to *Drosophila*, yeast and *Tetrahymena*, resulting in a longer left foot of the human 40S subunit (Fig. 3a and Supplementary Fig. 21). Conversely, ES9S is elongated in *Drosophila* and forms a 'horn' that interacts with S31e, thereby spanning the *Drosophila* 40S subunit region of the head comprising the binding site of eEF3 in yeast[32] (Fig. 3b and Supplementary Fig. 22). Although ES6S-A and ES6S-B are conserved in length between yeast, protists and metazoans, the conformations of these helices are markedly different between human and *Drosophila* (Fig. 3a, b), and between human, yeast and protists[4,6,8,10] (Supplementary Fig. 21). In addition, *Drosophila* ES6S-B contains a helical insertion resulting in branched ES6S-B1 and ES6S-B2 helices (Supplementary Fig. 21). Notably, the ES3S–ES6S region contributes to the binding site for the eukaryote-specific translation initiation factors eIF3 and eIF4G[33–35], emphasizing that structural variation in this region is likely to reflect functional differences during eukaryotic translation initiation.

Expansion segments of the human and *Drosophila* 60S subunit are mainly positioned on the side and back of the particle, with clusters located adjacent to the L1 and P stalks (Fig. 3c, d). Compared to yeast and protists, the most dramatic increase in mass is formed by ES7L, ES9L, ES10L, ES15L, ES27L and ES39L (Fig. 3c, d). Interestingly, the terminal loop of H30 within ES9L in the human rRNA forms continuous base pairs with an internal part of ES15L (Fig. 3c and Supplementary Fig. 23), analogous to the hybrid helix formed between ES3S–ES6S in the 40S subunit (Fig. 3a, b and Supplementary Fig. 21). The resulting mixed ES9L–ES15L helix seems to anchor the base of the human-specific extension of ES15L tightly to the surface of the particle.

Dynamic behaviour of expansion segments

As in yeast[10], human and *Drosophila* ES31L-A interacts with ribosomal protein S1e on the 40S subunit to form the eukaryote-specific intersubunit bridge eB8 (Fig. 4a–c). *Drosophila* ES31L is approximately 130 nucleotides longer than those of yeast and human (Supplementary Table 8), resulting in a prolonged helix ES31L-B that contacts L8e (Fig. 4a). Furthermore, helix ES31L-A is elongated and establishes a novel intersubunit bridge (which we term eB15, extending the nomenclature of yeast and protist ribosomes[10]) with ribosomal protein S27e near the mRNA exit site on the 40S subunit (Fig. 4a). In *Drosophila*, helix ES27L-C is extended compared to yeast ES27L, resulting in the formation of a second metazoan-specific intersubunit bridge (eB16) through interaction with S8e (Fig. 4a). Although human ES27L is larger than both those of yeast and *Drosophila*, contact to S8e is not observed in the human 80S ribosome because it adopts a conformation extending towards the L1 stalk (ES27L-in) (Fig. 4a–c). In addition to ES27L-in, an ES27L-out conformation that reaches towards

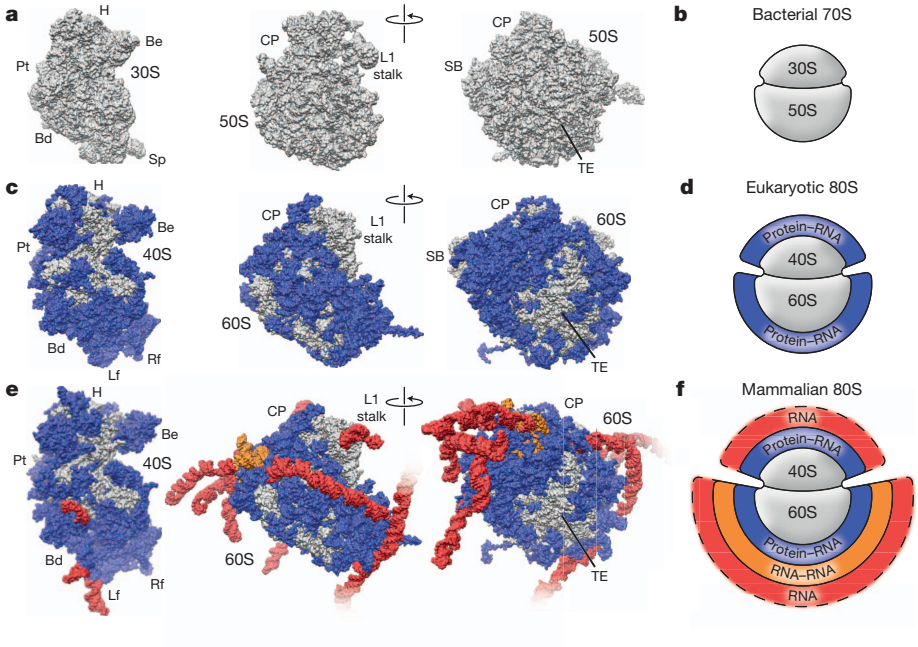


Figure 5 | Layered evolution of the eukaryotic ribosome. **a–f**, Surface representations (**a**, **c**, **e**) and schematics (**b**, **d**, **f**) of the bacterial *T. thermophilus* 70S ribosome (**a**, **b**)[47], the *S. cerevisiae* 80S ribosome (**c**, **d**)[10] (the eukaryote-specific protein–RNA layer is shown), and the mammalian 80S ribosome from *H. sapiens* (**e**, **f**) (the two additional layers, RNA–RNA and RNA-only, are shown). SB, P-stalk base; Sp, spur; TE, tunnel exit.

the tunnel exit has been observed previously for various yeast 80S ribosome complexes[6,36,37]. However, the *Drosophila* ES27L more closely resembles the intermediate ES27L position observed in wheat germ 80S ribosomes[6]. We therefore analysed the conformation of ES27L in the sub-populations of *Drosophila* and human 80S ribosomes that lacked eEF2 and exhibited non-rotated states. Although the ES27L-in conformation was identical between the eEF2-bound rotated and eEF2-lacking non-rotated human 80S ribosomes, a dynamic interplay of structural rearrangements was observed between *Drosophila* ES27L and ES31L (Fig. 4a, b): in the non-rotated state, we observe an ES27L-in conformation, such that the intersubunit bridge between ES27L-C and S8e is absent (Fig. 4b). In contrast, ES27L-B seems to displace ES31L-A to re-establish an intersubunit bridge with S27e (eB17) (Fig. 4b). Although ES31L-A maintains contact with S1e, the distal end of ES31L-A becomes disordered, presumably owing to the loss of interaction with S27e. The role of the dynamic rearrangements requires further investigation, but it seems that the conformational dynamics of ES27L and ES31L enable communication between two functional important regions of the ribosome, the mRNA exit site on the 40S subunit and the tunnel exit site on the 60S subunit. Indeed, deletion of ES27L in *Tetrahymena* is lethal[38], and the ES27L-out conformation has been observed to interact with a variety of important factors at the tunnel exit site, such as the nuclear export factor Arx1 (refs 39, 40), the ribosome-associated complex[41] and the membrane protein ERJ1 (Erj5p in *S. cerevisiae*)[42].

RNA–RNA interaction

It has been noted that ES31L and ES39L in yeast and *Tetrahymena* ribosomes use extended single-stranded (non-helical) rRNA stretches as platforms for interactions with ribosomal proteins[4,9,10]. In addition to ES31L and ES39L, the same structural principle is even more pronounced in metazoan ribosomes, and non-helical stretches are also observed in ES7L, ES10L and ES15L. Moreover, these structural elements are not only used for protein-RNA interactions but also establish RNA–RNA interactions between the expansion segments (Fig. 4d–g). In yeast and *Tetrahymena*, ES10L represents an asymmetric loop of 3 and 5 nucleotides, respectively, inserted into H38 (Fig. 4d, e). This non-helical insertion of ES10L has increased in *Drosophila* (by 12 nucleotides) and humans (by 22 nucleotides), leading to additional contacts with L30 and ES15L, respectively (Fig. 4f, g). In yeast, the loop of ES15L caps H45 and interacts with L4 and L18e (Fig. 4d), whereas in metazoans the insertion of helix ES15L-A creates an enlarged internal loop (Fig. 4f, g). In the *Drosophila* ribosome, the 9 non-helical nucleotides of this internal loop interact with ribosomal proteins L4, L18e and L28e, and also form contacts with the non-helical insertion of ES7L (Fig. 4f). The internal loop is further enlarged in the human ribosome, leading to new contacts with ribosomal proteins L6e and L30 as well as ES7L, ES9L and ES10L (Fig. 4g). Collectively, it seems that in metazoans, the non-helical insertions form a complex network of RNA–protein and RNA–RNA interactions that contribute to the stabilization of the large expansion segments cluster on the back of the 60S subunit.

Conclusion

The majority of the rRNA and ribosomal proteins that constitute the bacterial 70S ribosome is conserved in eukaryotes, and can therefore be considered to form the core of the 80S ribosome (Fig. 5a, b). Structures of the yeast and *Tetrahymena* ribosomes have revealed that the additional eukaryote-specific ribosomal proteins form a network of interactions with the rRNA expansion segments, resulting in an intertwined RNA–protein layer[6–10] (Fig. 5c, d). In metazoan eukaryotes, this RNA–protein layer has increased in size and complexity owing to the presence of additional ribosomal protein extensions and rRNA expansion-segment insertions (Fig. 5e, f and Supplementary Fig. 24). Moreover, the substantial increase in RNA mass of metazoans, particularly mammalian ribosomes, compared to yeast and protists, has resulted in the presence of two additional RNA layers (Fig. 5e, f): a rigid inner layer, resulting from multiple RNA–RNA tertiary interactions, followed by a flexible outer layer, arising from helical insertions and extensions of the rRNA expansion segments. The observed participation

of rRNA expansion segments in new intersubunit bridges or in the coordination of ribosomal ligands calls for further analysis of their functional significance in the complex environment of the eukaryotic cell.

METHODS SUMMARY

Drosophila and human 80S ribosomes were purified by sucrose density centrifugation from embryo extracts and peripheral blood mononuclear cells, respectively. For cryo-EM, ribosomes were vitrified and data were collected on a Titan Krios EM (FEI Company). Single-particle analysis and three-dimensional reconstruction were carried out using the SPIDER software package[43]. Ribosomal RNA was modelled using S2S[44] and Assemble[45]. Protein models were generated using Modeller[46].

Full Methods and any associated references are available in the online version of the paper.

Received 9 January; accepted 19 March 2013.

1. Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461,** 1234–1242 (2009).
2. Wilson, D. N. & Cate, J. H. D. The structure and function of the eukaryotic ribosome. *Cold Spring Harb. Perspect. Biol.* **4,** a011536 (2012).
3. Klinge, S., Voigts-Hoffmann, F., Leibundgut, M. & Ban, N. Atomic structures of the eukaryotic ribosome. *Trends Biochem. Sci.* **37,** 189–198 (2012).
4. Melnikov, S. *et al.* One core, two shells: bacterial and eukaryotic ribosomes. *Nature Struct. Mol. Biol.* **19,** 560–567 (2012).
5. Taylor, D. J. *et al.* Comprehensive molecular structure of the eukaryotic ribosome. *Structure* **17,** 1591–1604 (2009).
6. Armache, J. P. *et al.* Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. *Proc. Natl Acad. Sci. USA* **107,** 19748–19753 (2010).
7. Armache, J. P. *et al.* Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. *Proc. Natl Acad. Sci. USA* **107,** 19754–19759 (2010).
8. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331,** 730–736 (2011).
9. Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S. & Ban, N. Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* **334,** 941–948 (2011).
10. Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334,** 1524–1529 (2011).
11. Dube, P. *et al.* Correlation of the expansion segments in mammalian rRNA with the fine structure of the 80 S ribosome; a cryoelectron microscopic reconstruction of the rabbit reticulocyte ribosome at 21 A resolution. *J. Mol. Biol.* **279,** 403–421 (1998).
12. Spahn, C. M. *et al.* Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes; the IRES functions as an RNA-based translation factor. *Cell* **118,** 465–475 (2004).
13. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & Stark, H. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* **13,** 1695–1706 (2005).
14. Chandramouli, P. *et al.* Structure of the mammalian 80S ribosome at 8.7 A resolution. *Structure* **16,** 535–548 (2008).
15. Ruvinsky, I. & Meyuhas, O. Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. *Trends Biochem. Sci.* **31,** 342–348 (2006).
16. Koyama, Y., Katagiri, S., Hanai, S., Uchida, K. & Miwa, M. Poly(ADP-ribose) polymerase interacts with novel *Drosophila* ribosomal proteins, L22 and l23a, with unique histone-like amino-terminal extensions. *Gene* **226,** 339–345 (1999).
17. Ramakrishnan, V. Histone structure and the organization of the nucleosome. *Annu. Rev. Biophys. Biomol. Struct.* **26,** 83–112 (1997).
18. Taylor, D. J. *et al.* Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J.* **26,** 2421–2431 (2007).
19. Harms, J. M. *et al.* Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Mol. Cell* **30,** 26–38 (2008).
20. Gao, Y. G. *et al.* The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326,** 694–699 (2009).
21. Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **4,** a013706 (2012).
22. Ogle, J. M. & Ramakrishnan, V. Structural insights into translational fidelity. *Annu. Rev. Biochem.* **74,** 129–177 (2005).
23. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484,** 256–259 (2012).
24. Lu, H., Li, W., Noble, W. S., Payan, D. & Anderson, D. C. Riboproteomics of the hepatitis C virus internal ribosomal entry site. *J. Proteome Res.* **3,** 949–957 (2004).
25. Spahn, C. M. *et al.* Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291,** 1959–1962 (2001).
26. Balagopal, V. & Parker, R. Stm1 modulates translation after 80S formation in *Saccharomyces cerevisiae*. *RNA* **17,** 835–842 (2011).
27. Gerbi, S. A. in *Ribosomal RNA—Structure, Evolution, Processing, and Function in Protein Synthesis* (eds Zimmermann, R. A. & Dahlberg, A. E.) 71–87 (CRC Press, 1996).
28. Haga, J. Y., Hamilton, M. G. & Petermann, M. L. Electron microscopic observations on the large subunit of the rat liver ribosome. *J. Cell Biol.* **47,** 211–221 (1970).
29. Cannone, J. J. *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3,** 2 (2002).
30. Fields, D. S. & Gutell, R. R. An analysis of large rRNA sequences folded by a thermodynamic method. *Fold. Des.* **1,** 419–430 (1996).
31. Alkemar, G. & Nygard, O. Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. *Biochemistry* **45,** 8067–8078 (2006).
32. Andersen, C. B. *et al.* Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature* **443,** 663–668 (2006).
33. Srivastava, S., Verschoor, A. & Frank, J. Eukaryotic initiation factor-3 does not prevent association through physical blockage of the ribosomal subunit-subunit interface. *J. Mol. Biol.* **226,** 301–304 (1992).
34. Siridechadilok, B., Fraser, C. S., Hall, R. J., Doudna, J. A. & Nogales, E. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310,** 1513–1515 (2005).
35. Yu, Y., Abaeva, I. S., Marintchev, A., Pestova, T. V. & Hellen, C. U. Common conformational changes induced in type 2 picornavirus IRESs by cognate trans-acting factors. *Nucleic Acids Res.* **39,** 4851–4865 (2011).
36. Beckmann, R. *et al.* Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* **107,** 361–372 (2001).
37. Becker, T. *et al.* Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326,** 1369–1373 (2009).
38. Sweeney, R., Chen, L. H. & Yao, M. C. An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. *Mol. Cell. Biol.* **14,** 4203–4215 (1994).
39. Bradatsch, B. *et al.* Structure of the pre-60S ribosomal subunit with nuclear export factor Arx1 bound at the exit tunnel. *Nature Struct. Mol. Biol.* **19,** 1234–1241 (2012).
40. Greber, B. J., Boehringer, D., Montellese, C. & Ban, N. Cryo-EM structures of Arx1 and maturation factors Rei1 and Jjj1 bound to the 60S ribosomal subunit. *Nature Struct. Mol. Biol.* **19,** 1228–1233 (2012).
41. Leidig, C. *et al.* Structural characterization of a eukaryotic chaperone—the ribosome-associated complex. *Nature Struct. Mol. Biol.* **20,** 23–28 (2013).
42. Blau, M. *et al.* ERj1p uses a universal ribosomal adaptor site to coordinate the 80S ribosome at the membrane. *Nature Struct. Mol. Biol.* **12,** 1015–1016 (2005).
43. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
44. Jossinet, F. & Westhof, E. Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* **21,** 3320–3321 (2005).
45. Jossinet, F., Ludwig, T. E. & Westhof, E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26,** 2057–2059 (2010).
46. Eswar, N., Eramian, D., Webb, B., Shen, M. Y. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426,** 145–159 (2008).
47. Jenner, L., Demeshkina, N., Yusupova, G. & Yusupov, M. Structural rearrangements of the ribosome at the tRNA proofreading step. *Nature Struct. Mol. Biol.* **17,** 1072–1078 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Ungewickell for assistance with cryo-EM data collection and P. Palluch for preparation of peripheral blood mononucleic cells. We thank M. Yusupov, A. Ben-Shem, N. Garreau de Loubresse and S. Melnikov for sharing *S. cerevisiae* X-ray data before publication. We thank P. Becker for access to his fly facility and help with embryo collection, and V. Márquez, T. Fröhlich, G. Arnold, I. Forné and A. Imhof for mass-spectrometry analysis. This research was supported by grants from the Deutsche Forschungsgemeinschaft SFB594, SFB646 and GRK 1721 (to R.B.), and FOR1805 (to R.B. and D.N.W.). D.N.W. is supported by the European Molecular Biology Organization (EMBO) young investigator program. This work was supported by a European Research Council (ERC) Advanced Grant (to R.B.).

Author Contributions A.M.A. prepared *D. melanogaster* embryo extracts, purified *D. melanogaster* and *H. sapiens* ribosome samples, carried out mass-spectrometry analysis of *H. sapiens* ribosomes and prepared the figures; A.M.A. and J.-P.A. contributed blood, processed cryo-EM data and built atomic models; O.B. carried out cryo-EM data collection; M.H. performed deconvolution and sharpening on electron density maps; M.S. designed experiments for blood collection and peripheral-blood-mononuclear-cell preparations for human ribosome purification; A.M.A., J.-P.A., D.N.W. and R.B. interpreted results and wrote the manuscript. D.N.W. and R.B. designed research and supervised the project.

Author Information Coordinates of the atomic models have been deposited in the Protein Data Bank with accession numbers 3J38, 3J39, 3J3C and 3J3E for *Drosophila* 80S ribosomes and 3J3A, 3J3B, 3J3D and 3J3F for human 80S ribosomes. Full models can be obtained from the database of aligned ribosomal complexes (DARC) site (http://darcsite.genzentrum.lmu.de/darc/). Electron-microscopy maps of the *Drosophila* and human ribosomes have been deposited in the EM Data Bank under the accession codes EMD-5591 and EMD-5592, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. (beckmann@lmb.uni-muenchen.de).

METHODS

Purification of 80S ribosomes from *D. melanogaster*. Extracts from *D. melanogaster* embryos were prepared as described previously[48], and incubated under high-salt conditions (20 mM HEPES, pH 7.4, 500 mM potassium acetate (KOAc), 25 mM magnesium acetate ($Mg(OAc)_2$), 1 mM dithiotreitol (DTT), 0.5 mM phenylmethylsulfonylflourid (PMSF), 0.2 units per µl anti-RNase (Ambion)) with 0.5 mM puromycin for 15 min on ice, then for 10 min at 20 °C. Ribosomes were pelleted through a high-salt sucrose cushion (1 M sucrose, 500 mM KOAc, 25 mM $Mg(OAc)_2$, 1 mM DTT, 0.5 mM PMSF) at 355,040*g* (TLA120.2, Beckman-Coulter) for 60 min. The ribosomal pellet was suspended in buffer A (20 mM HEPES, pH 7.4, 100 mM KOAc, 5 mM $Mg(OAc)_2$, 1 mM DTT, 0.5 mM PMSF) with 125 mM sucrose. Ribosomes were purified further by centrifugation through a linear sucrose density gradient (10–40% sucrose in buffer A) at 202,048g (SW-40 Ti, Beckman Coulter) for 3 h at 4 °C. Fractions were collected using a Gradient Station (Biocomp) with an Econo UV Monitor (Biorad) and a FC203B Fraction Collector (Gilson). Ribosomes were pelleted from suitable fractions by centrifugation at 385,840*g* (TLA110, Beckman-Coulter) for 75 min. The pellet was suspended in buffer B (20 mM HEPES, pH 7.4, 50 mM KOAc, 2.5 mM $Mg(OAc)_2$, 100 mM sucrose, 1 mM DTT, 0.5 mM PMSF).

Purification of human 80S ribosomes. Mononuclear cells were prepared from human peripheral blood by ficoll-hypaque density-gradient centrifugation[49]. Cell pellets were suspended in lysis buffer (20 mM HEPES, pH 7.4, 100 mM KOAc, 7.5 mM $Mg(OAc)_2$, 1 mM DTT) with 1× Complete EDTA-free Protease Inhibitor cocktail (Roche) and lysed by repeated freeze and thaw cycles. Cell debris was removed by centrifugation for 20 min at 20,000*g* at 4 °C. Ribosomes were purified from the lysate after high salt and puromycin treatment as described above.

Electron microscopy and image processing. Samples were applied to 2-nm precoated R3/3 holey carbon supported copper grids (Quantifoil), vitrified using a Vitrobot Mark IV (FEI Company) and visualized on a Titan Krios transmission electron microscope (TEM) (FEI Company) under low-dose conditions (20e^- per Å^2) at a nominal magnification of ×75,000 with a nominal defocus between −1.0 and −3.5 µm. Data were collected using the semi-automated software EM-TOOLS (TVIPS GmbH) as described[50]. Contrast-transfer functions were determined using CTFFIND[51] and recorded images were manually inspected for good areas and power-spectra quality. Data were processed further using the SPIDER software package[43], in combination with an automated workflow as described previously[50].

The *D. melanogaster* 80S ribosome data set was collected at 300 keV at a magnification of ×128,200 at the plane of the charge-coupled device (CCD) using an Eagle 4k CCD camera (FEI Company, 4,096 × 4,096 pixels, 15-µm pixels, 5 s per full frame) resulting in an image pixel size of 1.17 Å on the object scale. The total data set consisted of 317,000 particles that entered a second round of selection using a machine-learning algorithm (MAPPOS[72]) that detects non-particles as described previously[50]. This procedure resulted in a cleaned data set of 288,000 (90.9%) particles that were used for the initial alignment. An empty yeast 80S ribosome structure was used as a reference. The data set was sorted[37,52] according to the presence of eEF2. The final (eEF2 and E-tRNA bound) data set contained 134,500 particles (42.4%) and reached a resolution of 6.0 Å after several rounds of refinement.

The *H. sapiens* 80S ribosome data set was collected at 200 keV at a magnification of ×148,721 at the plane of the CCD using a TemCam-F416 CMOS CCD camera (TVIPS GmbH, 4,096 × 4,096 pixels, 15.6-µm pixels, 1 s per full frame), resulting in a pixel size of 1.0489 Å on the object scale. Four separate data collections were used, of which the first (650,000 particles) was carried out using a normal field emission gun (FEG), whereas the remaining three (2.1-million particles) were collected with an X-FEG module (FEI Company) as the electron source. The collected data were initially aligned to a *Triticum aestivum* ribosome[6]. After a few rounds of refinement the data set was sorted[37,52], resulting in two maps representing stable conformations: a non-rotated 80S ribosome with E-tRNA, and a rotated 80S ribosome containing eEF2, SERPB1 and E-tRNA. The complete data were re-aligned using the best-resolved output from the previous refinement attempt (rotated 80S + eEF2 + SERPB1 + E-tRNA). After many rounds of refinement, re-sorting and application of a non-negative deconvolution and sharpening process[53], we arrived at a final average resolution of 5.4 Å from 343,343 particles. Local resolution was computed within a softened sphere (radius of 22 Å) at each voxel, as described previously[54], using the fourier shell correlation (FSC) of two reconstructions; from 50% of the particles and then the other 50%.

Ribosomal RNA modelling. *H. sapiens* 18S, 5S, 28S and 5.8S rRNA sequences were taken from GeneBank entries X03205 and V00589 and RefSeq accession numbers NR_003287 and NR_046235, respectively[55,56]. *D. melanogaster* sequences for the 18S, 28S, 2S, 5.8S and 5S rRNAs were obtained from GeneBank accessions M21017 and M25016, respectively[57,58], in combination with a revised 28S sequence for nucleotides 221–245 (H19 and H20), which are missing in the original sequence[59]. Structure-based sequence alignments of the conserved rRNA core were constructed using Sequence to Structure (S2S)[44] based on the X-ray structure of the 80S ribosome from *S. cerevisiae* (Protein Data Bank (PDB) accession codes 3O58 and 3O2Z)[10]. For the L1-stalk region (H76–H78) the corresponding structure of *Escherichia coli* (PDB accession 3R8S)[60] was used as template in a separate S2S alignment. All remaining parts of the rRNA were built *de novo* using Assemble[45], guided by features of the electron-density and secondary-structure predictions from RNAfold[61], in the main as described previously[6]. Secondary structures of large rRNA parts were predicted in smaller pieces and then by inspection of the corresponding electron-density map and subsequent model building. This generated new rRNA boundaries that were used as starting points for secondary-structure predictions of the following sequences. The iterative process resulted in the identification of simple, un-branched helical folds for the flexible human rRNA arms and enabled us to build complete molecular models of the human and *Drosophila* rRNA. The models were adjusted according to features of the electron density using Assemble[45], molecular dynamic flexible fitting (MDFF)[62] in visual molecular dynamics (VMD)[63] and Coot[64]. The reliability of the molecular model for the rRNA is indicated using the *b*-factor values within the PDB file; more reliably modelled regions have a lower *b*-factor.

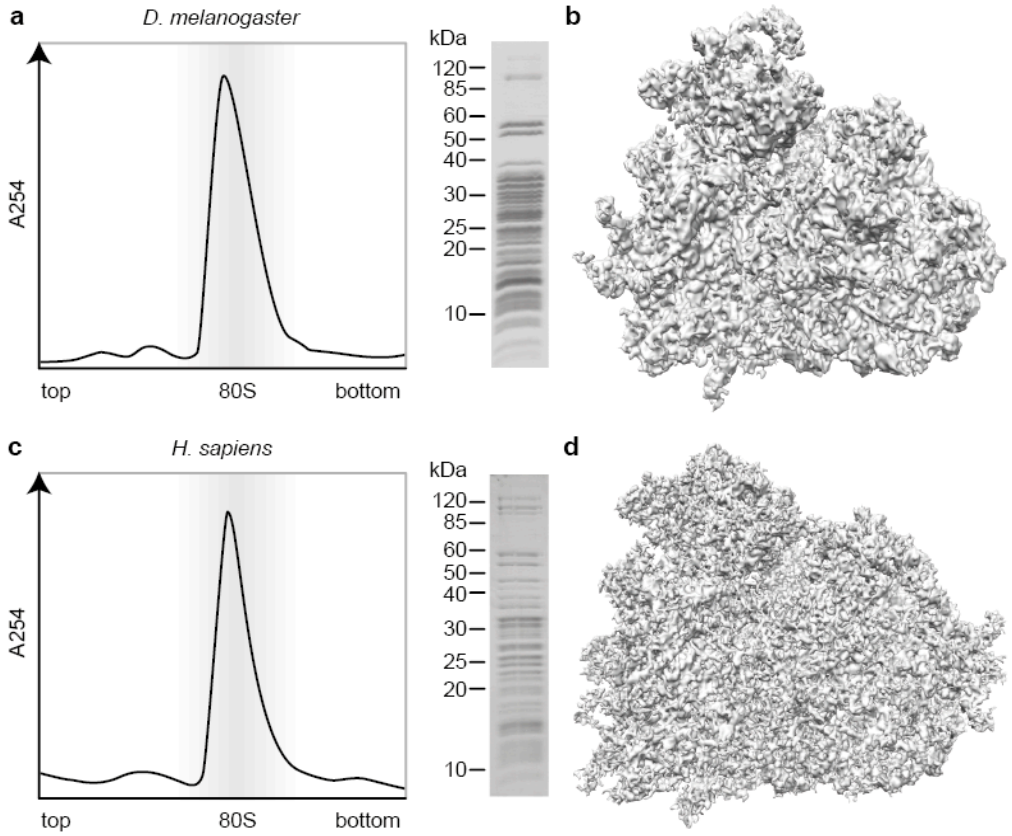
Ribosomal protein modelling. Owing to the availability of ribosomal 40S[8] and 60S structures[9] from *T. thermophila* and ribosomal 80S structures from *S. cerevisiae*[10], proteins were screened for the best fit into the *D. melanogaster* and *H. sapiens* densities. Protein multi-alignment was carried out using Jalview[65], ClustalW[66], ClustalOmega[67] and Mafft[68]. Results were extracted and Modeller[46] was used to create the initial models. Using UCSF Chimera[69] and Coot[64], they were fitted rigidly and adjusted into the densities. Subsequently, the remaining additional densities were analysed, assigned to specific protein entities and, in conjunction with secondary structure predictions, the models were extended. Furthermore, VMD[63], MDFF[62] and Coot were used to fix the clashes. The reliability of the molecular model for the ribosomal proteins is indicated using the *b*-factor values within the PDB file. More reliably modelled regions have a lower *b*-factor.

Mass-spectrometry analysis. For the *Drosophila* ribosome, proteins were extracted by acetic acid according to a previous study[70], and subsequent liquid chromatography tandem mass spectrometry (LC–MS/MS) analysis of the protein samples was carried out as described previously[71]. For the human ribosome sample, proteins were reduced, alkylated and digested with trypsin in solution before desalting and subsequent LC–MS/MS analysis using a LTQ Orbitrap XL (Thermo Scientific) mass spectrometer. MS/MS data were searched with Mascot (Matrix Science) using the SwissProt 2011.02 database and the following parameters: enzyme, trypsin; fixed modification, carbamidomethyl; variable modification, oxidation; peptide-mass tolerance, 10 p.p.m.; fragment mass tolerance, 0.8 Da; and up to two missed cleavages allowed.

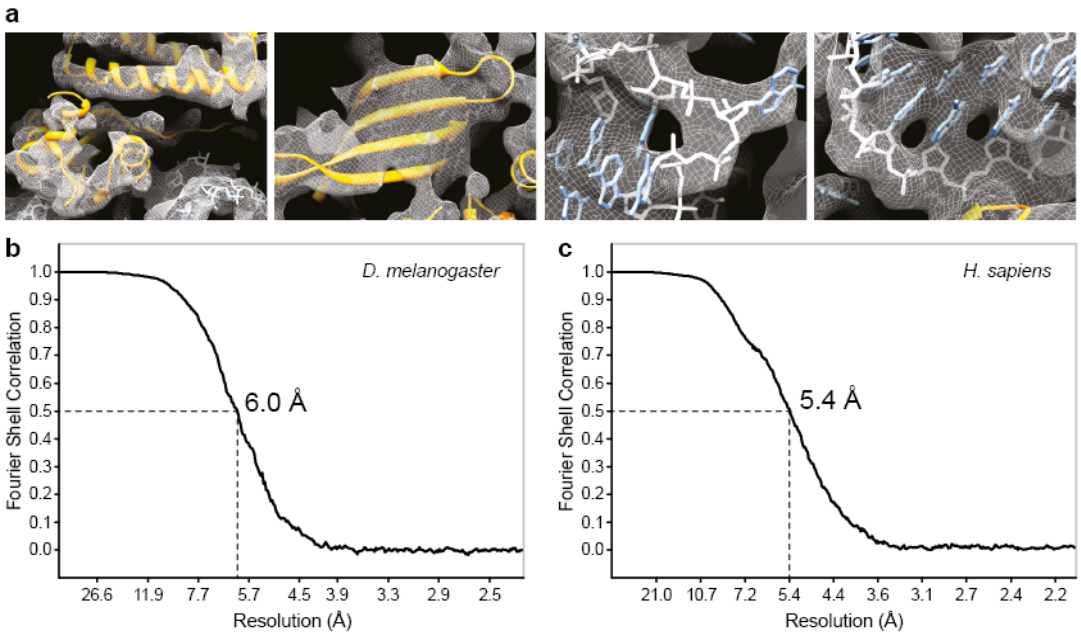
Figure preparation. Figures showing electron densities and atomic models were generated using UCSF Chimera[69].

48. Gebauer, F., Corona, D. F., Preiss, T., Becker, P. B. & Hentze, M. W. Translational control of dosage compensation in *Drosophila* by Sex-lethal: cooperative silencing via the 5′ and 3′ UTRs of msl-2 mRNA is independent of the poly(A) tail. *EMBO J.* **18,** 6146–6154 (1999).
49. Fuss, I. J., Kanof, M. E., Smith, P. D. & Zola, H. Isolation of whole mononuclear cells from peripheral blood and cord blood. *Curr. Protoc. Immunol.* **85,** 7.1.1–7.1.8 (2009).
50. Becker, T. *et al.* Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* **482,** 501–506 (2012).
51. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
52. Becker, T. *et al.* Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nature Struct. Mol. Biol.* **18,** 715–720 (2011).
53. Hirsch, M., Scholkopf, B. & Habeck, M. A blind deconvolution approach for improving the resolution of cryo-EM density maps. *J. Comput. Biol.* **18,** 335–346 (2011).
54. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl Acad. Sci. USA* **109,** 1380–1387 (2012).
55. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
56. Maden, B. E. *et al.* Clones of human ribosomal DNA containing the complete 18 S-rRNA and 28 S-rRNA genes. Characterization, a detailed map of the human ribosomal transcription unit and diversity among clones. *Biochem. J.* **246,** 519–527 (1987).
57. Tautz, D., Hancock, J. M., Webb, D. A., Tautz, C. & Dover, G. A. Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Mol. Biol. Evol.* **5,** 366–376 (1988).
58. Thompson, J. F., Wegnez, M. R. & Hearst, J. E. Determination of the secondary structure of *Drosophila melanogaster* 5 S RNA by hydroxymethyltrimethylpsoralen crosslinking. *J. Mol. Biol.* **147,** 417–436 (1981).
59. Rousset, F., Pelandakis, M. & Solignac, M. Evolution of compensatory substitutions through G.U intermediate state in *Drosophila* rRNA. *Proc. Natl Acad. Sci. USA* **88,** 10032–10036 (1991).
60. Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332,** 981–984 (2011).

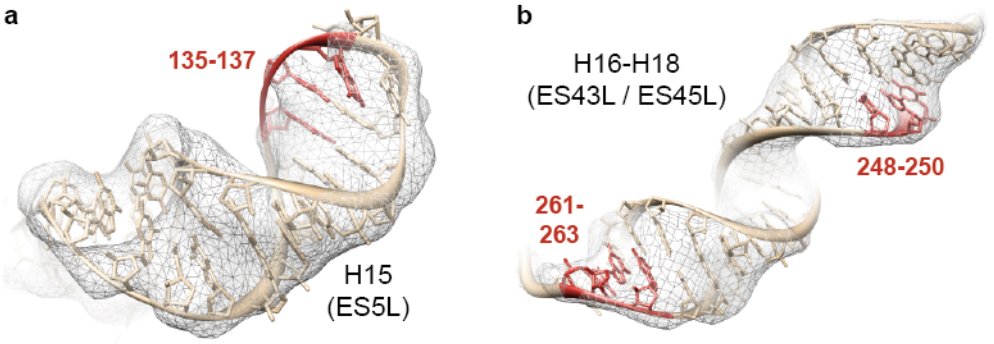
61. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31,** 3429–3431 (2003).
62. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16,** 673–683 (2008).
63. Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
64. Emsley, P. & Cowtan, K. Coot: model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).
65. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).
66. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
67. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
68. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33,** 511–518 (2005).
69. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
70. Nierhaus, K. H. & Dohme, F. Total reconstitution of functionally active 50S ribosomal subunits from *E. coli. Proc. Natl Acad. Sci. USA* **71,** 4713–4717 (1974).
71. Márquez, V. *et al.* Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J. Mol. Biol.* **405,** 1215–1232 (2011).
72. Norousi, R. *et al.* Automated post-picking using MAPPOS improves particle image detection from cryo-EM micrographs. *J. Struct. Biol.* http://dx.doi.org/10.1016/j.jsb.2013.02.008 (2013).



Supplementary Figure 1 | Purification and cryo-EM reconstructions of *Drosophila* and human 80S ribosomes. (**a**) Sucrose density gradient profile of the *D. melanogaster* ribosome preparation. The 80S peak fraction was analyzed by SDS-PAGE and coomassie staining (right). (**b**) Electron density of the *D. melanogaster* 80S ribosome deposited in the EM Data Bank (EMD-5591, filtered at 5 Å). (**c**) Sucrose density gradient profile of the *H. sapiens* ribosome preparation. The 80S peak fraction was analyzed by SDS-PAGE and coomassie staining (right). (**d**) Electron density of the *H. sapiens* 80S ribosome deposited in the EM Data Bank (EMD-5992, filtered at 4 Å).

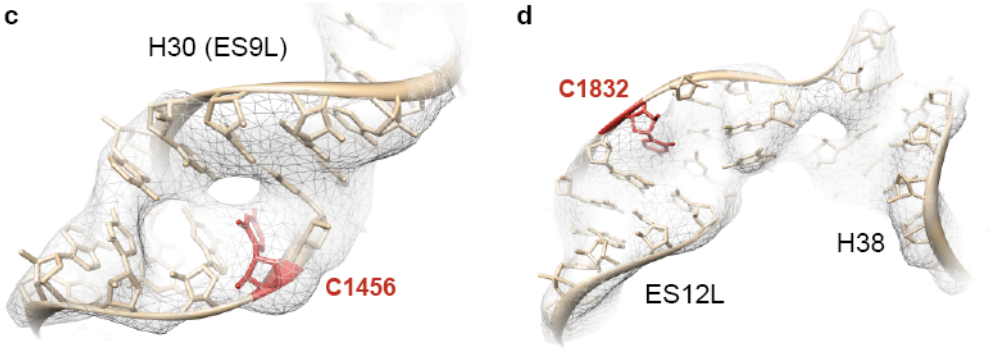


Supplementary Figure 2 | Electron density map features of the *Drosophila* 80S ribosome and resolution determination. (**a**) Selected views of the *Drosophila* 80S electron density map (gray mesh) and corresponding molecular models. Ribosomal proteins are shown in orange, rRNA in white (backbone) and blue (bases). (**b**, **c**) The 0.5 Fourier Shell Correlation cut-off criteria indicates that the cryo-EM maps of the (**b**) *D. melanogaster* and (**c**) *H. sapiens* 80S ribosome have average resolutions of 6.0 Å and 5.4 Å, respectively.



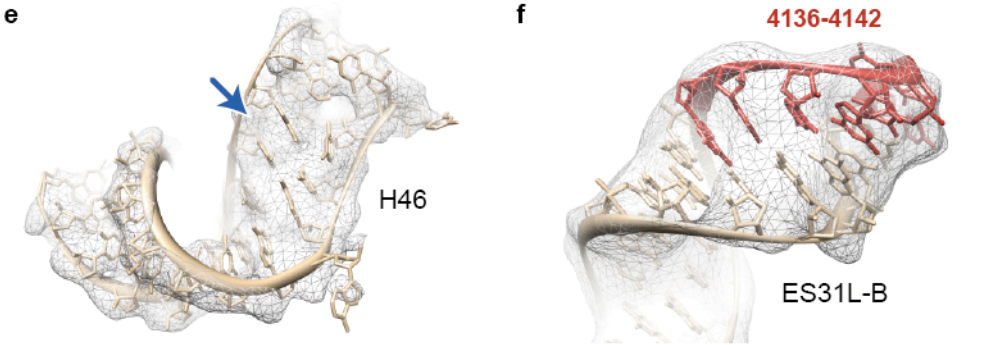

```
A)  126 CGCCCCGCGGCGGGGCGCGG 145
B)  126 CGCCCCGCG---GGGCGCGG 142
```

```
A)  246 GGCCCCCGGCGCGCCGGGCC 265
B)  243 GG---CCGGCGCGCG---CC 256
```




```
A) 1446 CCGCCGAGGGCGCACCACCG 1465
B) 1426 CCGCCGAGGG-GCACCACCG 1444
```

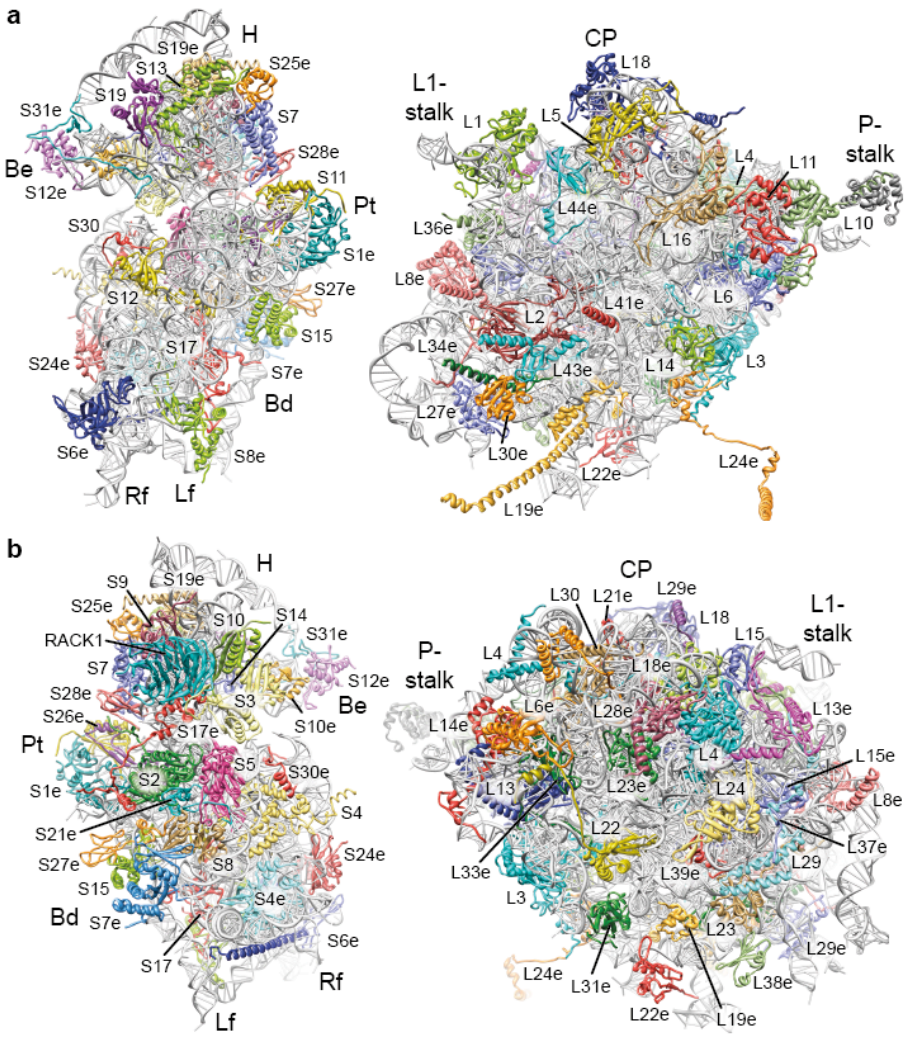
```
A) 1822 UGGAGCCGGGCGUGGAAUGC 1841
B) 1801 UGGAGCCGGG-GUGGAAUGC 1819
```



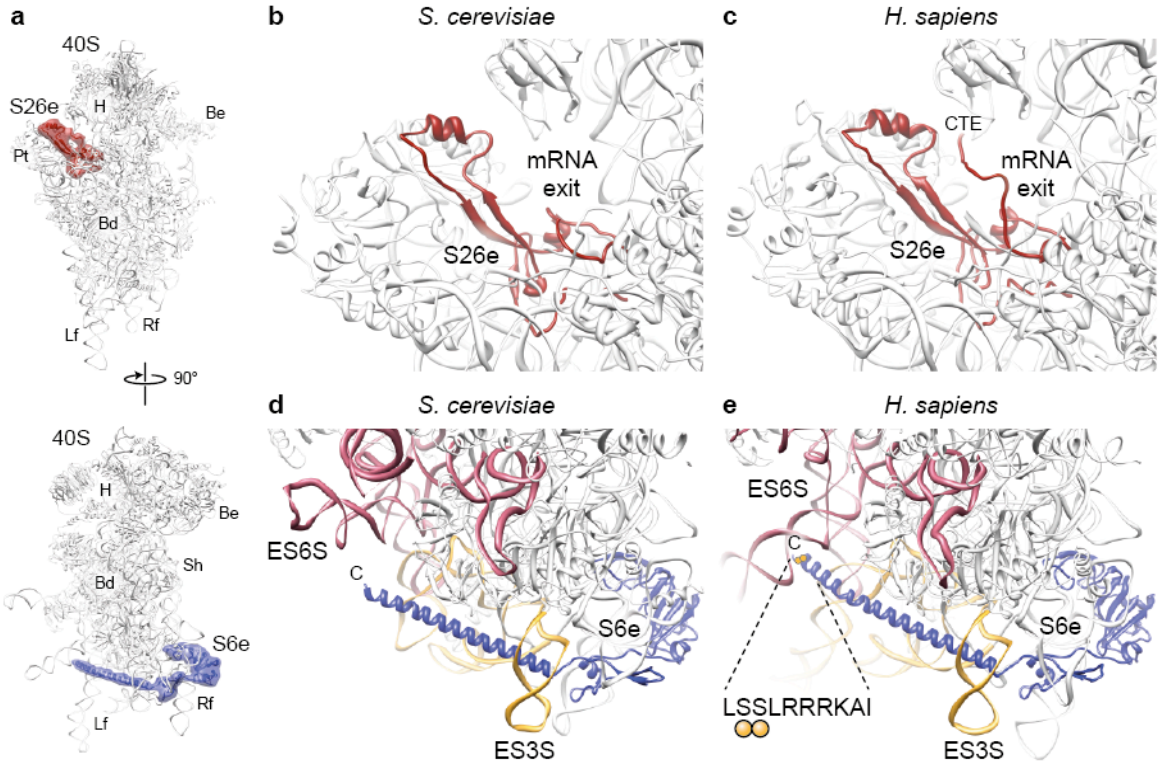

```
A) 2324 CCGGGUGGAG-CCGCCGCAG 2342
B) 2300 CCGGGUGGAGGCCGCCGCAG 2319
```

```
A) 4129 GCGCCCGGCCGCGCGCCGGC 4148
B) 4096 GCGCCCG-------CCCGGC 4108
```

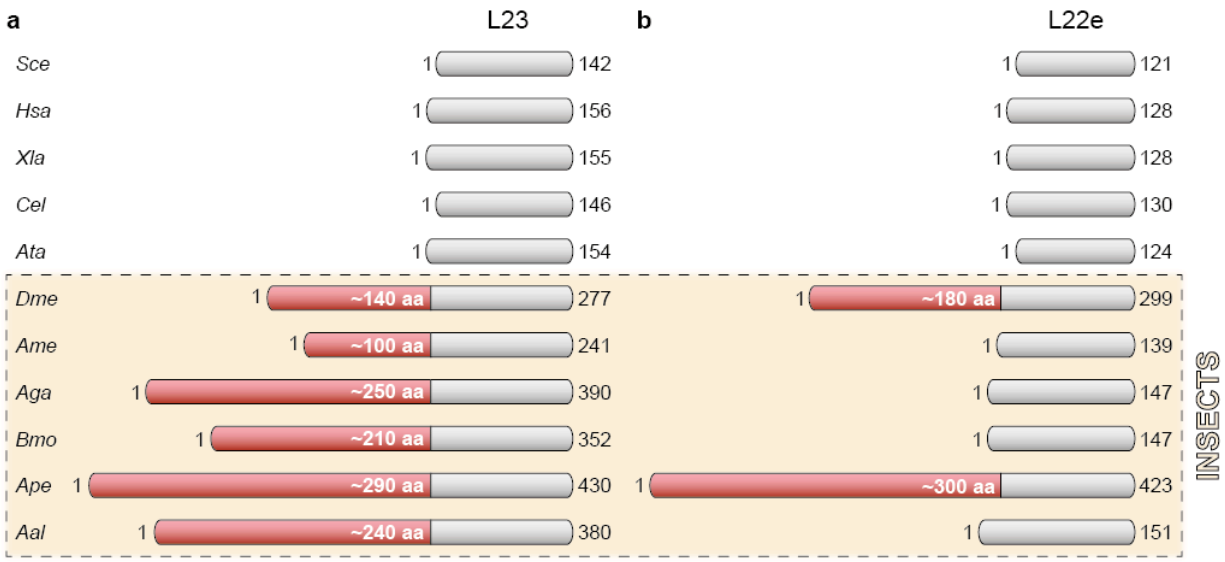
Supplementary Figure 3 | Cryo-EM allows distinction between variations in human 28S rRNA. (**a**) Electron density map (gray mesh) of H15 (ES5L) with corresponding molecular rRNA model (tan) based on the sequence of[55,56] (A). Nucleotides 135-137 (red) are clearly represented in the density but are absent in an alternative rRNA sequence[29,72] (B). (**b**) Similar representation of H16-H18 (ES43L and ES45L) as in (**a**). Two nucleotide patches, 248-250 and 261-263 (red) from rRNA sequence (A) are recognizable in the electron map. (**c**, **d**) Density map supports the presence of single nucleotide insertions C1456 and C1832 (red) within H30 (ES9L) and ES12L, respectively. (**e**) G2310 (highlighted blue in rRNA sequence (B)) is both not present in the electron density (blue arrow) and in the rRNA sequence (A) used to generate the molecular model. (**f**) Electron density of ES31L-B confirms the existence of nucleotides 4136-4142 (red). Map was filtered at 7 Å resolution for clarity.



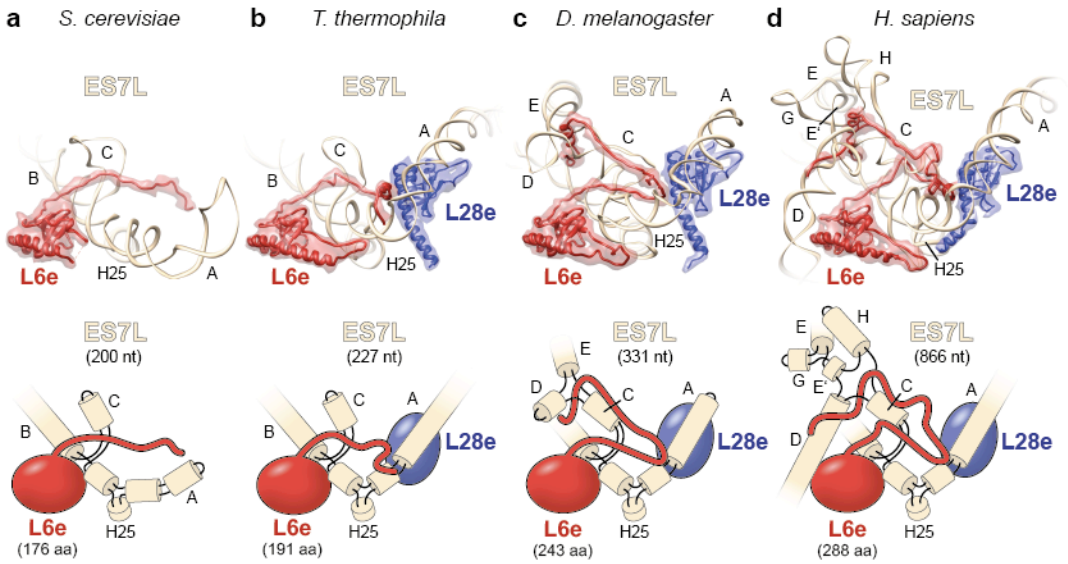
Supplementary Figure 4 | Protein architecture of the *D. melanogaster* 80S ribosome. (**a**) Interface (front) view of the 40S (left) and 60S (right) subunits. Ribosomal proteins are individually colored. Major landmarks are indicated: beak (Be), body (Bd), head (H), left foot (Lf), platform (Pt) and right foot (Rf) for the 40S, central protuberance (CP), L1-stalk and P-stalk for the 60S subunit. (**b**) Solvent-side (back) view of the 40S (left) and 60S (right) subunits. Proteins are colored as in (a).



Supplementary Figure 5 | Ribosomal proteins S26e and S6e. (**a**) Overviews showing the positions of ribosomal proteins S26e (top, red) and S6e (bottom, blue) on the 40S subunit. Major landmarks are labeled: beak (Be), body (Bd); head (H); left foot (Lf); platform (Pt); right foot (Rf). (**b**, **c**) Comparison of S26e from (**b**) *S. cerevisiae*[10] and (**c**) *H. sapiens*. The C-terminal extension of human S26e reaches into the mRNA exit channel. (**d**, **e**) Comparison of S6e from (**d**) *S. cerevisiae*[10] and (**e**) *H. sapiens*. High resolution structure of the human 80S ribosome allowed modeling of the helical C-terminal extension of S6e, including the conserved serine phosphorylation sites (indicated with orange spheres).



Supplementary Figure 6 | Schematic alignment of ribosomal proteins L23 and L22e. (**a**) Comparison of L23 from non-insect species (*Sce*, *S. cerevisiae*; *Hsa*, *H. sapiens*; *Xla*, *Xenopus laevis*; *Cel*, *Caenorhabditis elegans*; *Ata*, *Arabidopsis thaliana*) with the corresponding proteins from insects (boxed) (*Dme*, *D. melanogaster*; *Ame*, *Apis mellifera*; *Aga*, *Anopheles gambiae*; *Bmo*, *Bombyx mori*; *Ape*, *Acyrthosiphon pisum*; *Aal*, *Aedes albopictus*). (**b**) Comparison of L22e from non-insect species with the corresponding proteins from insects (boxed). Numbers indicate amino acid positions. The insect specific histone H1-like NTEs are highlighted in red.



Supplementary Figure 7 | Co-evolution of eukaryotic rRNA and r-proteins. (**a-d**) Interaction of r-proteins L6e and L28e with ES7L in (**a**) *S. cerevisiae* (lacking L28e)[6,7,10], (**b**) *Tetrahymena thermophila*[9], (**c**) *D. melanogaster* and (**d**) *H. sapiens* with corresponding schematic representations (bottom). In yeast, the NTE of L6e extends towards the position of L28e, which is lacking in yeast (a), but present in *Tetrahymena* (b) and all other higher eukaryotes (c, d). L28e interacts with ES7L-A, stabilizing and altering the position such that it interacts with NTE of L6e of *Tetrahymena*, *Drosophila* and human (b-d). The NTE of *Drosophila* L6e is elongated compared to lower eukaryotes, such as yeast and *Tetrahymena*, and clearly inserts through a three-way junction present in ES7L (c). The increase in size of human ES7L has led to the addition of helix ES7L-H converting the three-way junction into a four-way junction, as well as the addition of helices ES7L-E'/G to form a new three-way junction with ES7L-E (d). Moreover, compared to other higher eukaryotes, such as *Drosophila*, the NTE of mammalian L6e has a lysine-rich ~20 aa extension, which in the human 80S ribosome interacts with the mammalian-specific extended part of helix ES7L-D (d).

a *H. sapiens* L4

MACARPLISVYSEKGESSGKNVTLPAVFKAPIRPDIVNFVHTNLRKNNRQPYAVSELAGHQTSAESWGTGRAVARIPRVRGGGTHRSGQGAFGNMCRG
GRMFAPTKTWRRWHRRVNTTQKRYAICSALAASALPALVMSKGHRIEEVPELPLVVEDKVEGYKKTKEAVLLLKKLKAWNDIKKVYASQRMRAGKGKM
RNRRRIQRRGPCIIYNEDNGIIKAFRNIPGITLLNVSKLNILKLAPGGHVGRFCIWTESAFRKLDELYGTWRKAASLKSNYNLPMHKMINTDLSRILK
SPEIQRALRAPRKKIHRRVLKKNPLKNLRIMLKLNPYAKTMRRNTILRQARNHKLRVDKAAAAAAALQAKSDEKAAVAGKKPVVGKKGKKAAVGVKKQ
KKPLVGKKAAATKKPAPEKKPAEKKPTTEEKKPAA

b *H. sapiens* L14e

MVFRRFVEVGRVAYVSFGPHAGKLVAIVDVIDQNRALVDGPCTQVRRQAMPFKCMQLTDFILKFPHSAHQKYVRQAWQKADINTKWAATRWAKKIEAR
ERKAKMTDFDRFKVMKAKKMRNRIIKNEVKKLQKAALLKASPKKAPGTKGTAAAAAAAAAKVPAKKITAASKKAPAQKVPAQKATGQKAAPAPKAQK
GQKAPAQKAPAPKASGKKA

c *H. sapiens* L29e

MAKSKNHTTHNQSRKWHRNGIKKPRSQRYESLKGVDPKFLRNMRFAKKHNKKGLKKMQANNAKAMSARAEAIKALVKPKEVKPKIPKGVSRKLDRLAY
IAHPKLGKRARARIAKGLRLCRPKAKAKAKAKDQTKAQAAAPASVPAQAPKRTQAPTKASE

d *D. melanogaster* L22e

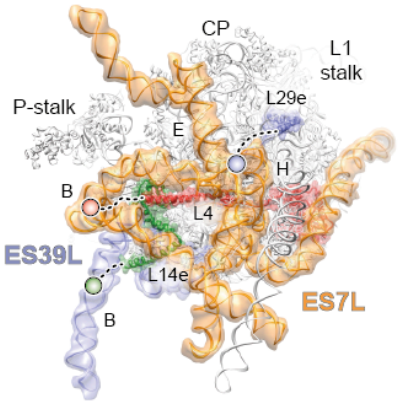
MAPTAKTNKGDTKTAAAKPAEKKAAPAAAAAKGKVEKPKAEAAKPAAAAAKNVKKASEAAKDVKAAAAAAKPAAAKPAAAKPAAASKDAGKKAPAAAA
PKKDAKAAAAPAPAKAAPAKKAASTPAAAPPAKKAAPAKAAAPAAAAPAPAAAAPAVAKPAPKPKAKAAPAPSKVVKKNVLRGKGQKKKKVSLRFTID
CTNIAEDSIMDVADFEKYIKARLKVNGKVNNLGNNVTFERSKLKLIVSSDVHFSKAYLKYLTKKYLKKNSLRDWIRVVANEKDSYELRYFRISSNDDE
DDDAE

e *D. melanogaster* L23

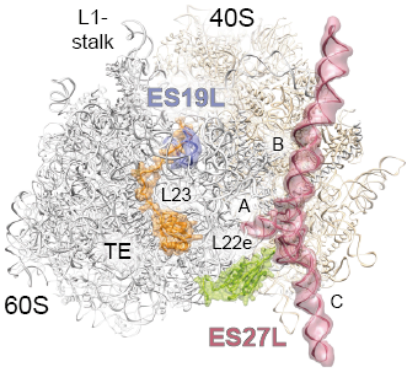
MPPKKPTEKSAKPGDKKPEQKKTAAAPAAGKKEAAPSAAKPAAAAPKKAAAPAAKKAAPAAKKPATAGAAAKKPAAVKTTAAAKAKSKDAKKKVLAGK
KPQSVLAKLSAKARAAAKAKKGVKPVTKPAKGTAKAKAVALLNAKKVQKKIIKGAFGTRARKIRTNVHFRRPTTLKLPRSPKYPRKSVPTRNRMDAYN
IIKYPLTTEAAMKKIEDNNTLVFLTHLRANKNHVRAAVRKLYDIKVAKVNVLIRPDGQKKAYVRLARDYDALDIANKIGII

---

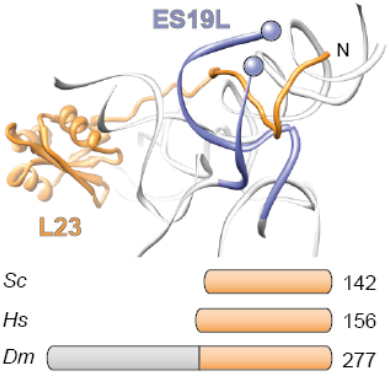
f *H. sapiens*



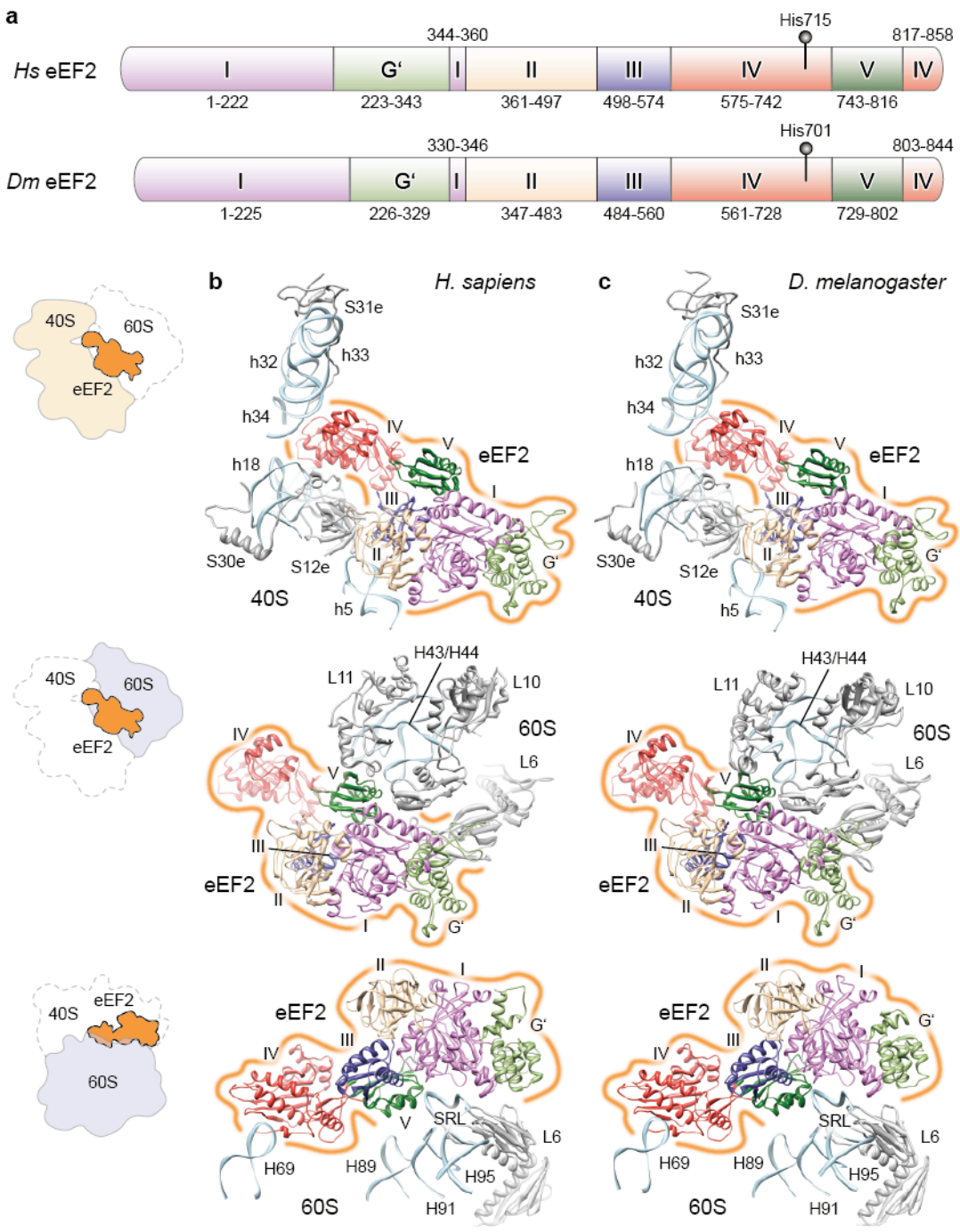
g *D. melanogaster*



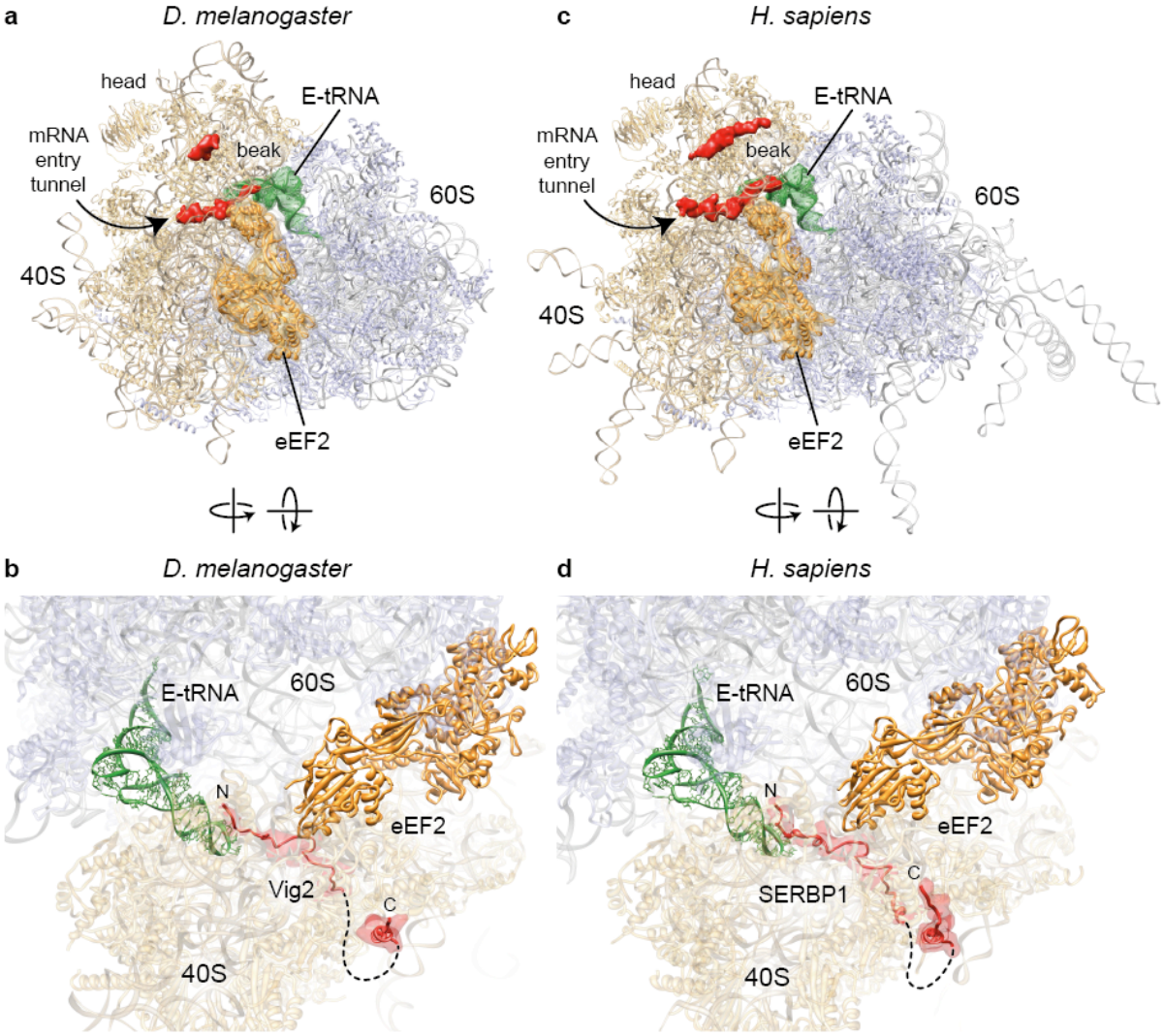
h *D. melanogaster*



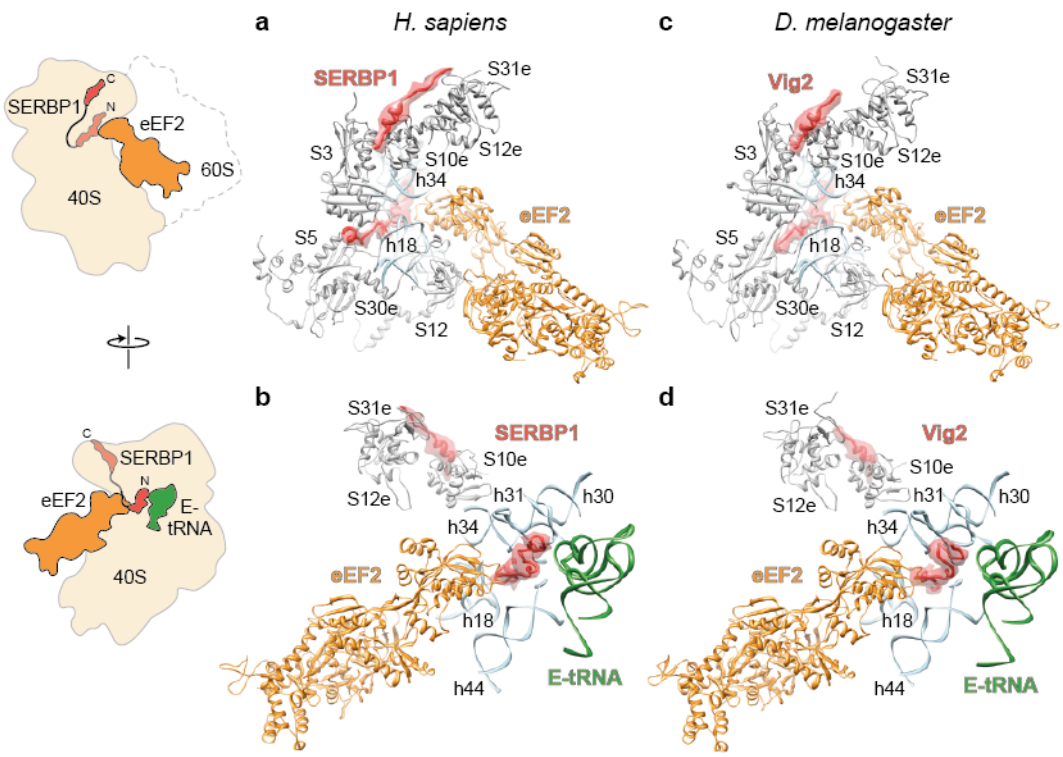
Supplementary Figure 8 | Histone H1-like extensions of r-proteins. (**a**-**e**) Sequences of human ribosomal proteins L4, L14e and L29e and *Drosophila* L22e and L23. A, K and P residues within the histone H1-like extensions are highlighted in red. (**f**) Back view of the human 60S subunit with r-proteins L4 (red), L14e (green) and L29e (blue) highlighted. Directions of the histone H1-like CTEs towards ES39L-B (violet) and ES7L-B and ES7L-E/H (orange) are indicated with dashed lines and spheres. (**g**) Location of *Drosophila* r-proteins L22e (green) and L23 (orange) with respect to ES27L (dark purple) and ES19L (violet). The NTE of L22e approaches the 3-way junction created by ES27L-A/B/C, while the NTE of L23 is located directly adjacent to the insect-specific 28S rRNA excision site within ES19L[73], indicating a possible involvement of L23 in rRNA processing. (**h**) Close-up view of *Drosophila* L23 (orange) in direct vicinity of ES19L (violet) (top). Blue spheres mark the insect-specific processing sites within ES19L. Schematic alignment of L23 from *S. cerevisiae* (*Sc*), *H. sapiens* (*Hs*) and *D. melanogaster* (*Dm*) (bottom). Numbers indicate amino acid positions.


a
Hs eEF2
Dm eEF2
His715
His701
344-360
330-346
817-858
803-844
I
G'
II
III
IV
V
1-222
223-343
361-497
498-574
575-742
743-816
1-225
226-329
347-483
484-560
561-728
729-802
40S
60S
eEF2
b
H. sapiens
c
D. melanogaster
S31e
h32
h33
h34
h18
S30e
S12e
h5
H43/H44
L11
L10
L6
SRL
H69
H89
H95
H91

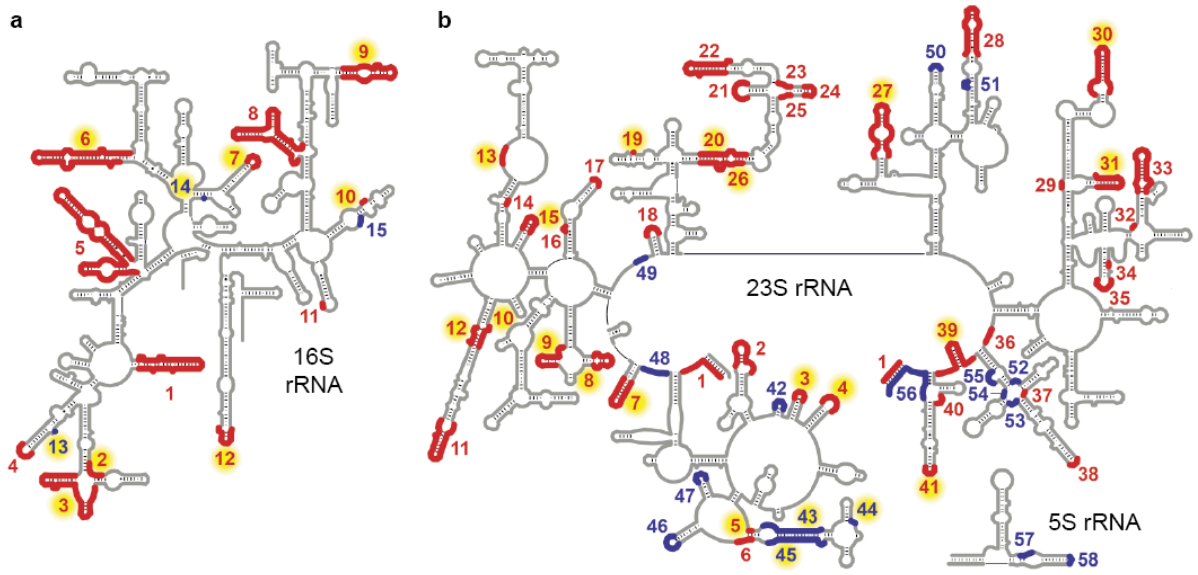
Supplementary Figure 9 | Interaction of eEF2 with the ribosome. (**a**) Schematic representation of *H. sapiens* (*Hs*) and *D. melanogaster* (*Dm*) eEF2 showing domain organization (I to V and G', individually colored) of the proteins. Numbers indicate domain boundaries. The conserved histidine residues (His701 and His715 in *D. melanogaster* and *H. sapiens*, respectively) that are target for modification to diphthamide are highlighted. (**b**, **c**) Interactions of (**b**) human and (**c**) *Drosophila* eEF2 with the ribosome (r-proteins and rRNA colored in gray and light blue, respectively). Domains of eEF2 are colored as in (a). The overall shape of eEF2 is indicated with an orange line. Views are illustrated by small insets (left) with eEF2 colored in orange.


a
D. melanogaster
head
E-tRNA
mRNA
entry
tunnel
beak
60S
40S
eEF2
c
H. sapiens
head
E-tRNA
mRNA
entry
tunnel
beak
60S
40S
eEF2
b
D. melanogaster
E-tRNA
60S
N
eEF2
Vig2
C
40S
d
H. sapiens
E-tRNA
60S
N
eEF2
SERBP1
C
40S

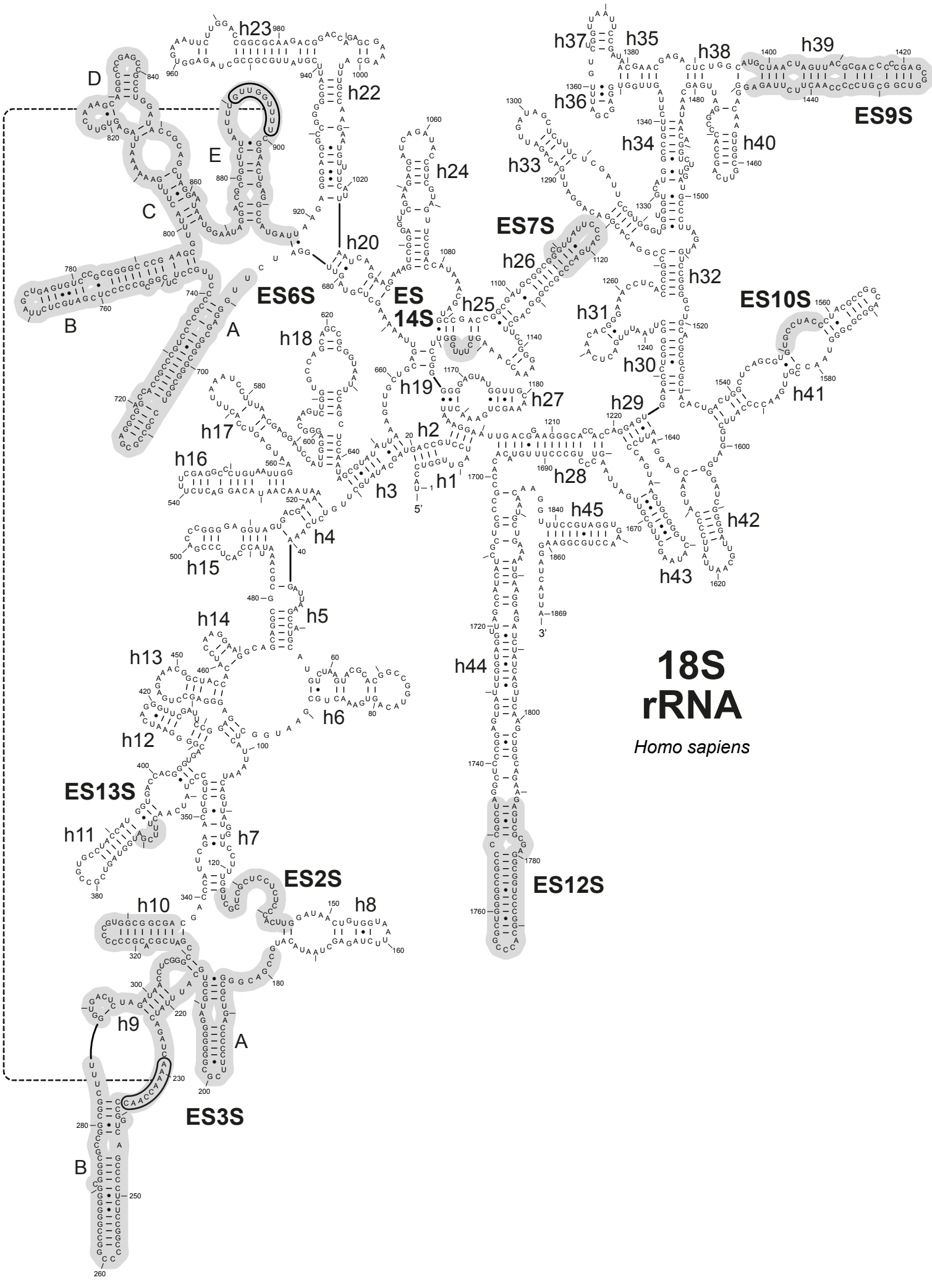
Supplementary Figure 10 | SERBP1, Vig2, E-site tRNA and eEF2 bound to the ribosome. (**a**) *D. melanogaster* 80S ribosome (40S and 60S subunits in light orange and light blue, respectively) with bound E-site tRNA (green) and eEF2 (orange). Experimental electron density that was assigned to Vig2 (after subtracting all other modeled parts from the structure) is highlighted in red. (**b**) Top view of the 40S and 60S subunits of *D. melanogaster* with eEF2 (orange) and tRNA (green) bound to the A- and E-site, respectively. Vig2 (red) binds along the mRNA path until the P-site. (**c**) *H. sapiens* 80S ribosome (40S and 60S subunits in light orange and light blue, respectively) with bound E-site tRNA (green) and eEF2 (orange). Experimental electron density that was assigned to SERBP1 (after subtracting all other modeled parts from the structure) is highlighted in red. (**d**) Top view of the 40S and 60S subunits of *H. sapiens* with eEF2 (orange) and tRNA (green) bound to the A- and E-site, respectively. SERBP1 (red) binds along the mRNA path until the P-site.



Supplementary Figure 11 | Interaction of human SERBP1 and *Drosophila* Vig2 with the ribosome. (**a**, **b**) Interaction of human SERBP1 (red) with eEF2 (orange), E-tRNA (green) and the 40S ribosomal subunit (r-proteins and rRNA colored in gray and light blue, respectively). (**c**, **d**) Interaction of *Drosophila* Vig2 (red) with eEF2 (orange), E-tRNA (green) and the 40S ribosomal subunit (r-proteins and rRNA colored in gray and light blue, respectively). Views are indicated by insets (left). SERBP1 and Vig2 are shown together with transparent electron density generated from the models to highlight the structures.



Supplementary Figure 12 | Secondary structure diagram of bacterial rRNA indicating sites of variable regions and eukaryotic ES. (**a**) 16S rRNA secondary structure diagram from *E. coli* with variable regions (VR) defined by Gerbi[27] colored and numbered in red. Newly defined VR are colored and numbered in blue. VR that correspond to ES in eukaryotes are highlighted in yellow (**b**) 23S/5.8S/5S rRNA secondary structure diagram from *E. coli* with variable regions colored as in (a) (see Supplementary Table 7 and 8 for exact VR/ES definitions and more detailed information on the extended ES nomenclature).


18S
rRNA
Homo sapiens
h1
h2
h3
h4
h5
h6
h7
h8
h9
h10
h11
h12
h13
h14
h15
h16
h17
h18
h19
h20
h22
h23
h24
h25
h26
h27
h28
h29
h30
h31
h32
h33
h34
h35
h36
h37
h38
h39
h40
h41
h42
h43
h44
h45
ES6S
ES14S
ES7S
ES9S
ES10S
ES13S
ES2S
ES12S
ES3S
A
B
C
D
E
5'
3'

Supplementary Figure 13 | Secondary structure diagram of the *H. sapiens* 18S rRNA. The diagram was taken from the Comparative RNA Web (CRW) Site (www.rna.ccbb.utexas.edu)[29] and modified according to the final rRNA model. Nucleotides, helices and ES are numbered. Canonical base pairs are depicted with (-), while (•) denote GU wobble base pairs. Gray regions indicate rRNA ES.

28S / 5.8S rRNA

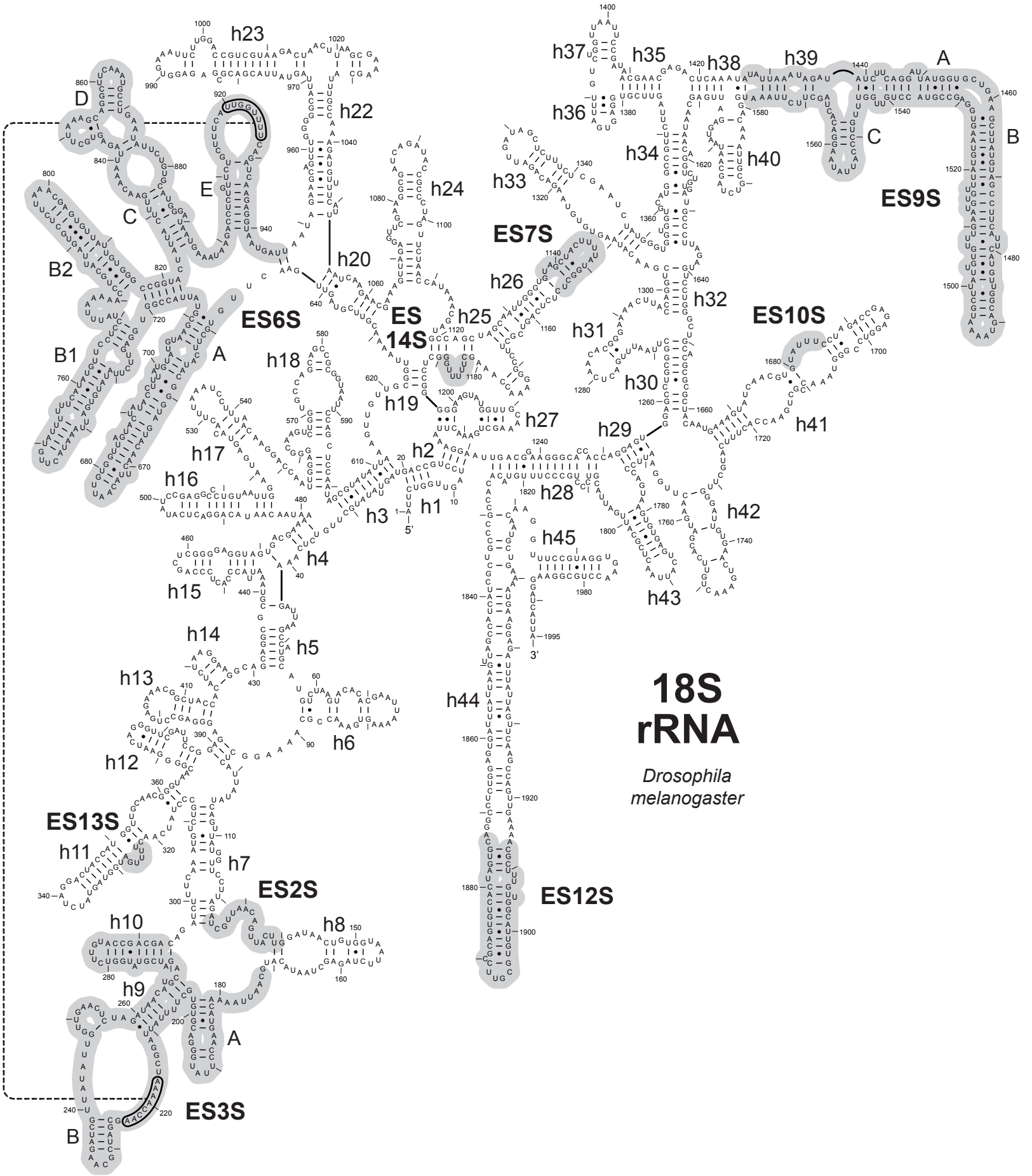
Homo sapiens

5S rRNA

Homo sapiens

ES15L
ES27L
ES30L
ES31L
ES19L
ES20L
ES26L
ES13L
ES12L
ES10L
ES8L
ES3L
ES4L
ES7L
ES5L
ES43L
ES44L
ES45L
ES41L
ES9L
ES39L

Supplementary Figure 14 | Secondary structure diagrams of the *H. sapiens* 5S/5.8S/28S rRNAs. The 5S and 5.8S/28S rRNA diagrams were taken from the CRW Site[29] and[30], respectively. The latter are accessible via the CRW Site. All diagrams were modified according to the final rRNA model and include nucleotide, helix and ES numbering. Canonical base pairs are depicted with (-), while (•) denote GU wobble base pairs. Gray regions indicate rRNA ES.


18S
rRNA
Drosophila
melanogaster
ES6S
ES
14S
ES7S
ES9S
ES10S
ES13S
ES2S
ES12S
ES3S
h1
h2
h3
h4
h5
h6
h7
h8
h9
h10
h11
h12
h13
h14
h15
h16
h17
h18
h19
h20
h22
h23
h24
h25
h26
h27
h28
h29
h30
h31
h32
h33
h34
h35
h36
h37
h38
h39
h40
h41
h42
h43
h44
h45
A
B
B1
B2
C
D
E
5'
3'

Supplementary Figure 15 | Secondary structure diagram of the *D. melanogaster* 18S rRNA. The diagram was taken from the CRW Site[29] and modified according to the final rRNA model. Nucleotides, helices and ES are numbered. Canonical base pairs are depicted with (-), while (•) denote GU wobble base pairs. Gray regions indicate rRNA ES.

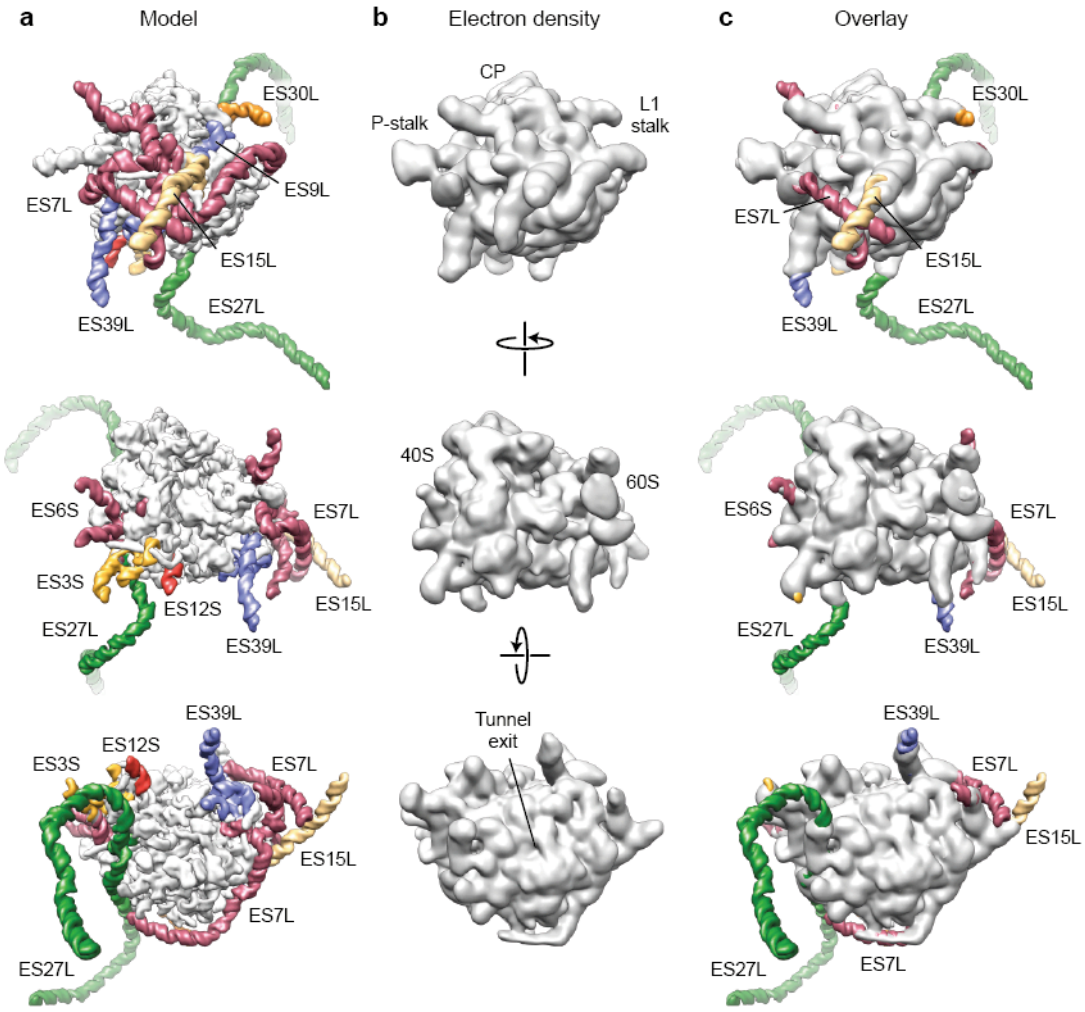
28S / 5.8S / 2S rRNA

Drosophila melanogaster

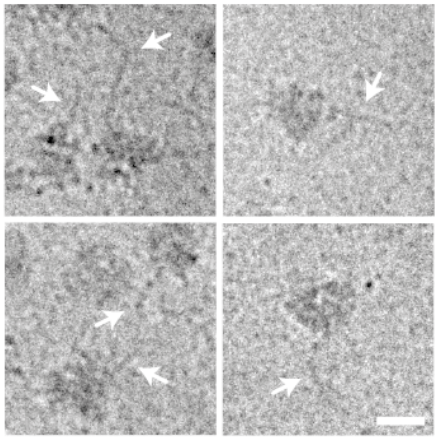
5S rRNA

Drosophila melanogaster

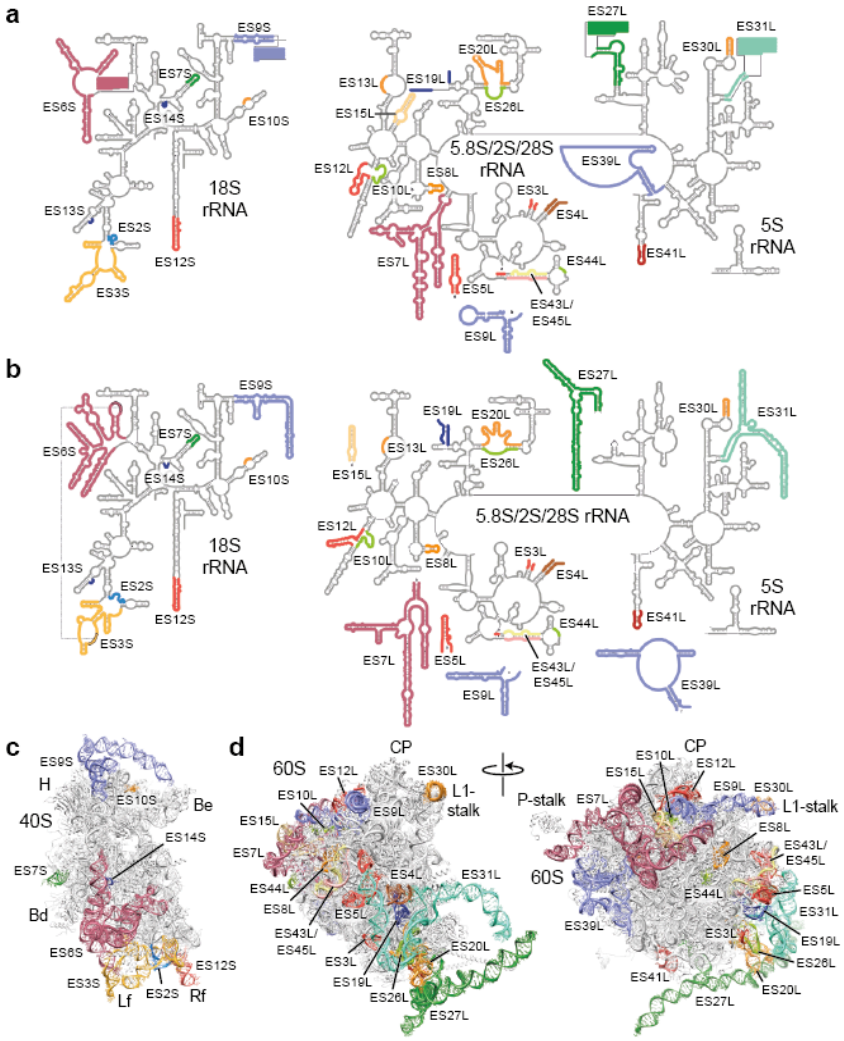
Supplementary Figure 16 | Secondary structure diagrams of the *D. melanogaster* 2S/5S/5.8S/28S rRNAs. The 5S and 2S/5.8S/28S rRNA diagrams were taken from the CRW Site[29] and[30], respectively. The latter are accessible via the CRW Site. All diagrams were modified according to the final rRNA model and include nucleotide, helix and ES numbering. Canonical base pairs are depicted with (-), while (•) denote GU wobble base pairs. Gray regions indicate rRNA ES.



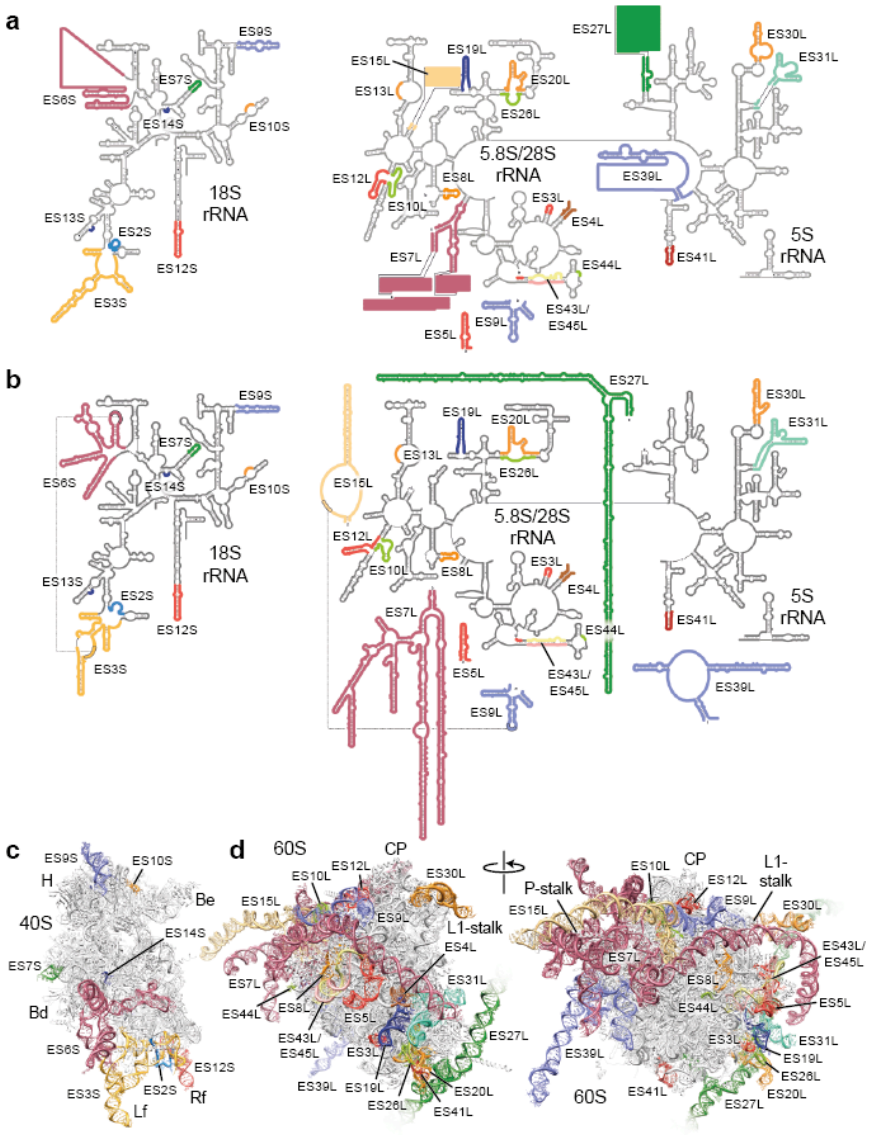
Supplementary Figure 17 | Lack of density for distal portion of ES. (**a**) Model of the human 80S ribosome converted into electron density and filtered at 8 Å in three different views. ES are colored distinctly. (**b**) Experimental electron density of the human 80S ribosome filtered at 20 Å to visualize flexible parts of the structure. The density map is shown in the same views as in (a). (**c**) Overlay of (a) and (b). The experimental density includes all ES core structures, leaving only few distal parts unsupported due to their highly flexible nature (ES3S, ES6S, ES7L, ES15L, ES30L and ES39L). Modeling of these parts was guided by secondary structure predictions that indicate predominantly unbranched, helical folds. Due to lack of electron density, ES27L is placed in an arbitrary position.



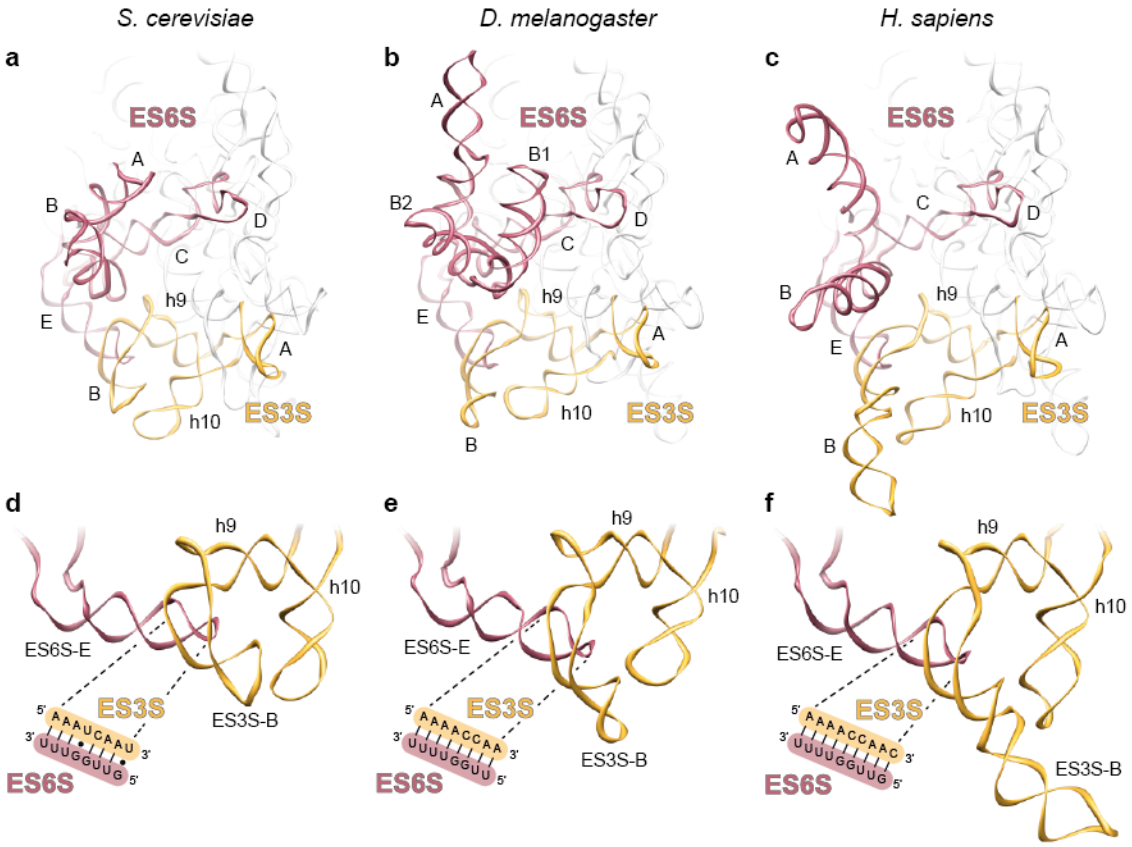
Supplementary Figure 18 | Cryo-EM images of the human 80S ribosome. Micrographs of vitrified human 80S ribosomes were taken under low dose conditions at a nominal magnification of 90,000 × on a Tecnai G2 Spirit (FEI Company) at 120 keV. Note the extended rRNA tentacles (arrows). Scale bar is 20 nm.



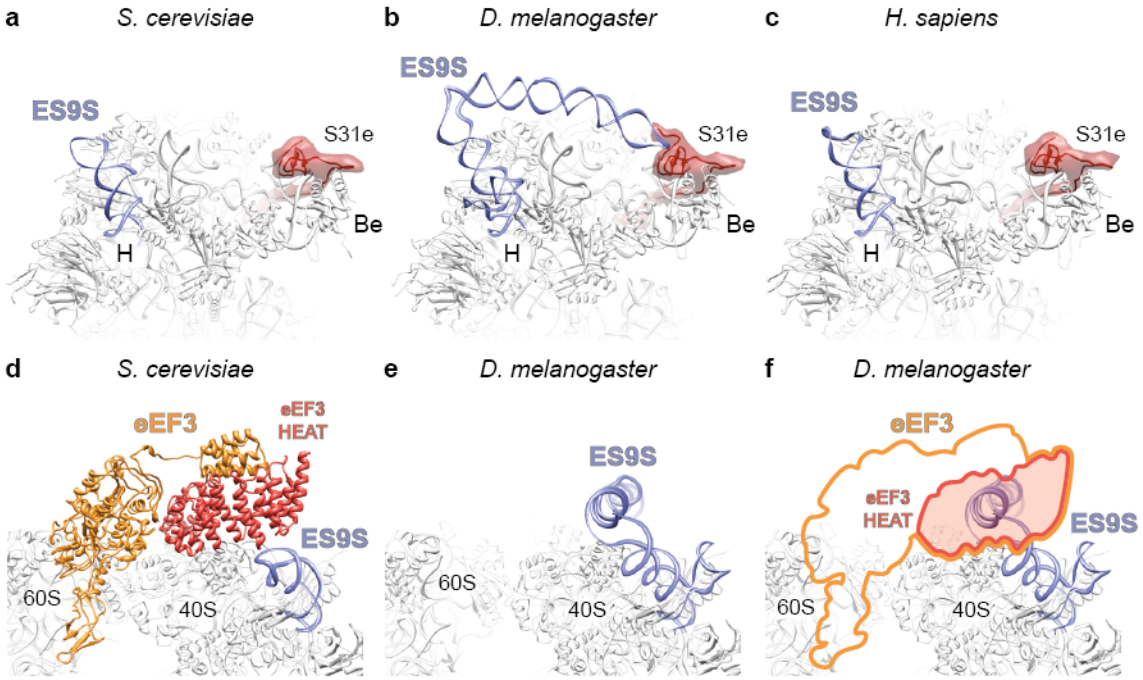
Supplementary Figure 19 | Revised version of the *D. melanogaster* rRNA secondary structure diagram. (**a**) Schematic representation of the original *D. melanogaster* rRNA secondary structure diagram as obtained from[29,30] with ES colored distinctly. The secondary structure was not predicted for several ES parts (colored boxes and extended single stranded regions). (**b**) Complete revised version of the *D. melanogaster* rRNA secondary structure. ES colored as in (a). (**c-d**) Molecular models of the (**c**) 40S and (**d**) 60S subunits from *D. melanogaster* with ES colored as in (a, b). Landmarks include the beak (Be), body (Bd), head (H) left foot (Lf), and right foot (Rf) for the small subunit and central protuberance (CP), L1-stalk, and P-stalk for the large subunit.


a
ES9S
ES7S
ES6S
ES14S
ES10S
18S
rRNA
ES13S
ES2S
ES12S
ES3S
ES27L
ES30L
ES31L
ES19L
ES15L
ES20L
ES13L
ES26L
5.8S/28S
rRNA
ES12L
ES8L
ES39L
ES10L
ES3L
ES4L
ES7L
ES44L
ES41L
5S
rRNA
ES43L/
ES45L
ES5L
ES9L
b
ES9S
ES7S
ES6S
ES14S
ES10S
18S
rRNA
ES13S
ES2S
ES12S
ES3S
ES27L
ES30L
ES31L
ES19L
ES20L
ES13L
ES26L
ES15L
5.8S/28S
rRNA
ES12L
ES10L
ES8L
ES3L
ES4L
ES7L
ES44L
ES41L
5S
rRNA
ES5L
ES43L/
ES45L
ES39L
ES9L
c
ES9S
ES10S
H
Be
40S
ES14S
ES7S
Bd
ES6S
ES12S
ES3S
ES2S
Rf
Lf
d
60S
CP
ES10L
ES12L
ES30L
ES15L
ES9L
L1-stalk
ES4L
ES7L
ES31L
ES44L
ES8L
ES5L
ES43L/
ES45L
ES27L
ES3L
ES39L
ES19L
ES20L
ES26L
ES41L
P-stalk
ES10L
CP
ES12L
L1-
stalk
ES15L
ES9L
ES30L
ES43L/
ES45L
ES7L
ES8L
ES44L
ES5L
ES3L
ES31L
ES39L
ES19L
ES26L
60S
ES41L
ES20L
ES27L

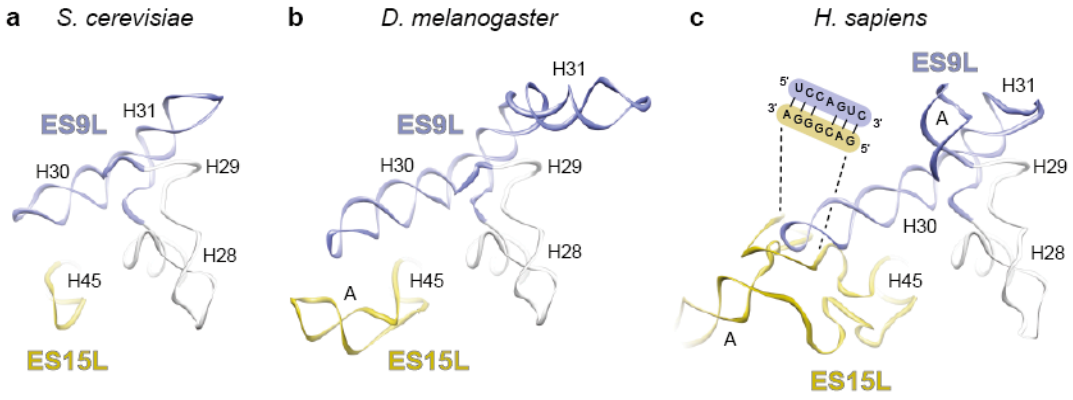
Supplementary Figure 20 | Revised version of the *H. sapiens* rRNA secondary structure diagram. (**a**) Schematic representation of the original *H. sapiens* rRNA secondary structure diagram as obtained from[29,30] with ES colored distinctly. The secondary structure was not predicted for several ES parts (colored boxes and extended single stranded regions). (**b**) Complete revised version of the *H. sapiens* rRNA secondary structure. ES colored as in (a). (**c**-**d**) Molecular models of the (**c**) 40S and (**d**) 60S subunits from *H. sapiens* with ES colored as in (a, b). Landmarks include the beak (Be), body (Bd), head (H) left foot (Lf), and right foot (Rf) for the small subunit and central protuberance (CP), L1-stalk, and P-stalk for the large subunit.



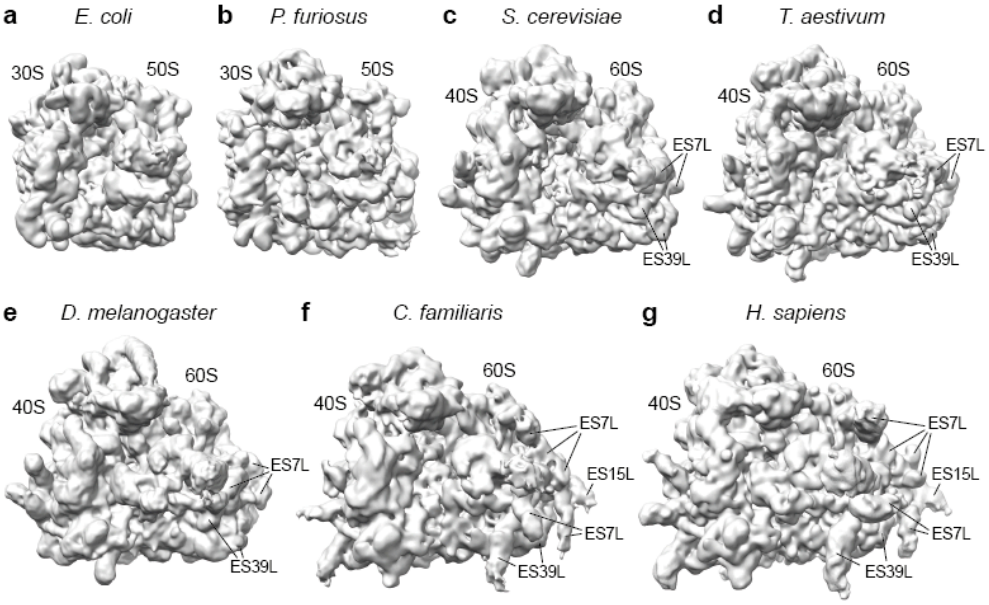
Supplementary Figure 21 | Interaction of ES3S and ES6S on the small ribosomal subunit. (**a-c**) Comparison of ES3S (bright orange) and ES6S (dark purple) between (**a**) *S. cerevisiae*[10], (**b**) *D. melanogaster* and (**c**) *H. sapiens*. (**d-f**) Conserved interaction of ES3S and ES6S via a hybrid helix between ES6S-E and an internal loop of ES3S-B in (**d**) *S. cerevisiae*[10], (**e**) *D. melanogaster* and (**f**) *H. sapiens*. Colors are used as in (a). Nucleotides involved in hybrid helix formation are shown in secondary structure drawings. Canonical base pairs are depicted with (-), while (•) denote GU wobble base pairs.



Supplementary Figure 22 | Interaction of *Drosophila* ES9S with S31e and binding site of eEF3 on the 40S subunit. (**a-c**) ES9S (violet) and ribosomal protein S31e (red) in (**a**) *S. cerevisiae*[10], (**b**) *D. melanogaster* and (**c**) *H. sapiens*. In *Drosophila* the tip of the extended ES9S-B helix interacts with S31e and thereby links the head (H) and beak (Be) regions of the small ribosomal subunit. (**d**) Head region of the 40S subunit from *S. cerevisiae*[10] with bound eEF3 (orange/red) taken from[32]. ES9S is colored in violet and the HEAT domain of eEF3 highlighted in red. (**e**) Head region of the 40S subunit from *D. melanogaster* with ES9S colored in violet. (**f**) Overlay of (e) with a schematic outline of eEF3 from (d). The HEAT domain of eEF3 would sterically clash with the extended *Drosophila* ES9S. Whereas eEF3 is fungal-specific, the translation regulator GCN1 has homology to the ribosome binding HEAT domain of eEF3[74], and is thus likely to interact with the same region in higher eukaryotic ribosomes.



Supplementary Figure 23 | Interaction of ES9L and ES15L on the large ribosomal subunit. (**a-c**) Comparison of ES9L (violet) and ES15L (green) between (**a**) *S. cerevisiae*[10], (**b**) *D. melanogaster* and (**c**) *H. sapiens*. Human ES9L and ES15L interact via a hybrid helix between the tip of H30 and an extended internal loop of ES15L. Nucleotides involved in hybrid helix formation are shown in secondary structure drawing. Canonical base pairs are depicted with (-).



Supplementary Figure 24 | Gallery of ribosome structures determined by cryo-EM. (**a-g**) Comparison of cryo-EM maps of (**a**) bacterial *E. coli* 70S[75] and (**b**) archaeal *Pyrococcus furiosus* 70S[76] ribosomes with eukaryotic 80S ribosomes from (**c**) yeast (*S. cerevisiae*)[6,7], (**d**) wheat-germ (*T. aestivum*)[6,7], (**e**) fruitfly (*D. melanogaster*), (**f**) dog (*Canis familiaris*)[14] and (**g**) human *H. sapiens*. All maps were filtered at the same resolution (10 Å) for comparison. Densities corresponding to ES7L and ES39L are indicated.

Supplementary Table 1 | *H. sapiens* 80S ribosome MS data.

Large subunit proteins				Small subunit proteins			
Protein name	Old human name	Uniprot ID	Score	Protein name	Old human name	Uniprot ID	Score
L1	L10A	P62906	766	RACK1	RACK1	P63244	297
L2	L8	P62917	2043	S1e	S3A	P61247	1144
L3	L3	P39023	1540	S2	SA	P08865	757
L4	L4	P36578	2267	S3	S3	P23396	771
L5	L11	P62913	921	S4	S9	P46781	138
L6	L9	P32969	1023	S4e	S4	P62701	919
L6e	L6	Q02878	812	S5	S2	P15880	1117
L8e	L7A	P62424	1610	S6e	S6	P62753	1277
L10	LP0	P05388	605	S7	S5	P46782	301
L11	L12	P30050	789	S7e	S7	P62081	1054
L13	L13A	P40429	872	S8	S15A	P62244	139
L13e	L13	P26373	931	S8e	S8	P62241	1126
L14	L23	P62829	1001	S9	S16	P62249	695
L14e	L14	P50914	1057	S10	S20	P60866	756
L15	L27A	P46776	1109	S10e	S10	P46783	753
L15e	L15	P61313	421	S11	S14	P62263	854
L16	L10	P27635	1150	S12	S23	P62266	878
L18	L5	P46777	937	S12e	S12	P25398	239
L18e	L18	Q07020	1078	S13	S18	P62269	432
L19e	L19	P84098	1116	S14	S29	P62273	482
L20e	L18A	Q02543	940	S15	S13	P62277	294
L21e	L21	P46778	977	S17	S11	P62280	763
L22	L17	P18621	829	S17e	S17	P08708	820
L22e	L22	P35268	158	S19	S15	P62841	417
L23	L23A	P62750	292	S19e	S19	P39019	1037
L24	L26	P61254	483	S21e	S21	P63220	604
L24e	L24	P83731	870	S24e	S24	P62847	263
L27e	L27	P61353	448	S25e	S25	P62851	494
L28e	L28	P46779	657	S26e	S26	P62854	349
L29	L35	P42766	210	S27e	S27	P42677	136
L29e	L29	P47914	814	S28e	S28	P62857	170
L30	L7	P18124	766	S30e	S30	P62861	144
L30e	L30	P62888	530	S31e	S27A	P62979	386
L31e	L31	P62899	520				
L32e	L32	P62910	604				
L33e	L35A	P18077	199				
L34e	L34	P49207	460				
L36e	L36	Q9Y3U8	651				
L37e	L37	P61927	59				
L38e	L38	P63173	272				
L39e	L39	P62891	48				
L40e	L40	P62987	131				
L41e	L41	P62945	n.d.	Ribosome associated proteins			
L43e	L37A	P61513	402				
L44e	L36A	P83881	301	Protein name		Uniprot ID	Score
P1	LP1	P05386	185	eEF2		P13639	2654
P2	LP2	P05387	286	SERBP1		Q8NC51	212

The Mascot score, highlighted in green is given for identified proteins. Not detected (n.d.) proteins are indicated in red.

Supplementary Table 2 | *D. melanogaster* 80S ribosome MS data.

Large subunit proteins				Small subunit proteins			
Protein name	Old *Drosophila* name	Uniprot ID	Score	Protein name	*Old Drosophila* name	Uniprot ID	Score
L1	L10A	Q9VTP4	1043	RACK1	RACK1	O18640	n.d.
L2	L8	Q9V3G1	1691	S1e	S3A	P55830	865
L3	L3	O16797	2797	S2	SA	P38979	118
L4	L4	P09180	1389	S3	S3	Q06559	850
L5	L11	P46222	385	S4	S9	P55935	264
L6	L9	P50882	310	S4e	S4	P41042	824
L6e	L6	Q9V9W2	799	S5	S2	P31009	546
L8e	L7A	P46223	956	S6e	S6	P29327	712
L10	LP0	P19889	620	S7	S5	Q24186	253
L11	L12	Q9W1B9	611	S7e	S7	Q9VA91	537
L13	L13A	Q9VNE9	370	S8	S15A	P48149	390
L13e	L13	P41126	601	S8e	S8	Q8MLY8	1284
L14	L17A	P48159	1018	S9	S16	Q9W237	245
L14e	L14	P55841	887	S10	S20	P55828	429
L15	L27A	P41092	507	S10e	S10	Q9VB14	497
L15e	L15	O17445	60	S11	S14	P14130	898
L16	L10	O61231	622	S12	S23	Q8T3U2	186
L18	L5	Q9W5R8	1279	S12e	S12	P80455	872
L18e	L18	Q9VS34	609	S13	S18	P41094	570
L19e	L19	P36241	662	S14	S29	Q9VH69	271
L20e	L18A	P41093	206	S15	S13	Q03334	355
L21e	L21	Q9V9M7	613	S17	S11	Q0E9B6	406
L22	L17	Q9W3W8	404	S17e	S17	P17704	585
L22e	L22	P50887	1219	S19	S15	Q7JZW2	516
L23	L23A	Q9W0A8	888	S19e	S19	P39018	305
L24	L26	Q9VVU2	762	S21e	S21	O76927	437
L24e	L24	Q9VJY6	580	S24e	S24	Q9W229	520
L27e	L27	Q9VBN5	335	S25e	S25	P48588	345
L28e	L28	Q9VZS5	592	S26e	S26	P13008	192
L29	L35	Q9W499	n.d.	S27e	S27	Q9VBU9	256
L29e	L29	Q24154	219	S28e	S28	Q9W334	94
L30	L7	P32100	629	S30e	S30	Q9VDH8	n.d.
L30e	L30	Q9VJ19	481	S31e	S27A	P15357	409
L31e	L31	Q9V597	314				
L32e	L32	P04359	416				
L33e	L35A	Q9VNB9	n.d.				
L34e	L34	Q9VBH8	94				
L36e	L36	P49630	213				
L37e	L37	Q9VXX8	106				
L38e	L38	Q9W5N2	224				
L39e	L39	O16130	n.d.				
L40e	L40	P18101	n.d.				
L41e	L41	Q962S2	n.d.				
L43e	L37A	Q9VMU4	368				
L44e	L36A	Q9VLT7	260				
P1	LP1	P08570	617				
P2	LP2	P05389	55				

Ribosome associated proteins		
Protein name	Uniprot ID	Score
eEF2	P13060	522
Vig2	Q960D3	418

The Mascot score, highlighted in green is given for identified proteins. Not detected (n.d.) proteins are indicated in red.

Supplementary Table 3 | Small subunit, eEF2 and SERBP1 models from *H. sapiens*.

Protein name	Protein family	Range of sequence	Modeled sequence	Uniprot_ID	Template used	Template organism
RACK1	RACK1	1-317	2-314	P63244	3U5C_g	*Saccharomyces cerevisiae*
S1e	S3ae	1-264	19-233	P61247	3U5G_B	*Saccharomyces cerevisiae*
S2	S2p	1-295	2-209	P08865	3U5G_A	*Saccharomyces cerevisiae*
S3	S3p	1-243	1-227	P23396	3U5C_D	*Saccharomyces cerevisiae*
S4	S4p	1-194	7-188	P46781	3U5G_J	*Saccharomyces cerevisiae*
S4e	S4e	1-263	1-263	P62701	3U5G_E	*Saccharomyces cerevisiae*
S5	S5p	1-293	53-278	P15880	3U5G_C	*Saccharomyces cerevisiae*
S6e	S6e	1-249	1-237	P62753	3U5C_G	*Saccharomyces cerevisiae*
S7	S7p	1-204	14-204	P46782	3U5G_F	*Saccharomyces cerevisiae*
S7e	S7e	1-194	5-194	P62081	3U5C_H	*Saccharomyces cerevisiae*
S8	S8p	1-130	2-130	P62244	3U5C_W	*Saccharomyces cerevisiae*
S8e	S8e	1-208	2-207	P62241	3U5C_I	*Saccharomyces cerevisiae*
S9	S9p	1-146	6-146	P62249	3U5C_Q	*Saccharomyces cerevisiae*
S10	S10p	1-119	16-119	P60866	3U5G_U	*Saccharomyces cerevisiae*
S10e	S10e	1-165	1-98	P46783	3U5C_K	*Saccharomyces cerevisiae*
S11	S11p	1-151	16-151	P62263	3U5C_O	*Saccharomyces cerevisiae*
S12	S12p	1-143	1-142	P62266	3U5G_X	*Saccharomyces cerevisiae*
S12e	S12e	1-132	9-132	P25398	3U5C_M	*Saccharomyces cerevisiae*
S13	S13p	1-152	6-142	P62269	3U5G_S	*Saccharomyces cerevisiae*
S14	S14p	1-56	4-56	P62273	3U5C_d	*Saccharomyces cerevisiae*
S15	S15p	1-151	2-151	P62277	3U5C_N	*Saccharomyces cerevisiae*
S17	S17p	1-158	1-158	P62280	3U5C_L	*Saccharomyces cerevisiae*
S17e	S17e	1-135	1-126	P08708	3U5G_R	*Saccharomyces cerevisiae*
S19	S19p	1-145	4-130	P62841	3U5G_P	*Saccharomyces cerevisiae*
S19e	S19e	1-145	4-144	P39019	3U5G_T	*Saccharomyces cerevisiae*
S21e	S21e	1-83	1-82	P63220	3U5G_V	*Saccharomyces cerevisiae*
S24e	S24e	1-133	3-128	P62847	3U5C_Y	*Saccharomyces cerevisiae*
S25e	S25e	1-125	41-115	P62851	3U5G_Z	*Saccharomyces cerevisiae*
S26e	S26e	1-115	2-108	P62854	3U5G_a	*Saccharomyces cerevisiae*
S27e	S27e	1-84	1-84	P42677	3U5C_b	*Saccharomyces cerevisiae*
S28e	S28e	1-69	5-68	P62857	3U5G_c	*Saccharomyces cerevisiae*
S30e	S30e	1-59	1-59	P62861	3U5C_e	*Saccharomyces cerevisiae*
S31e	S27ae	77-156	82-152	P62979	3U5C_f	*Saccharomyces cerevisiae*
eEF2	eEF2	1-858	3-858	P13639	1U2R_A 2XQD_Z 2NPF_T	*Saccharomyces cerevisiae* *Saccharomyces cerevisiae* *Thermus thermophilus*
Serpine1	PAIRB	1-408	139-188, 281-303	Q8NC51	3U5C_h	*Saccharomyces cerevisiae*

Supplementary Table 4 | Large subunit protein models from *H. sapiens*.

Protein name	Protein family	Range of sequence	Modeled sequence	Uniprot_ID	Template used	Template organism
L1	L1p	1-217	1-217	P62906	2HW8_A	*Thermus thermophilus*
L2	L2p	1-257	2-256	P62917	3U5E_A	*Saccharomyces cerevisiae*
L3	L3p	1-403	2-398	P39023	3U5E_B	*Saccharomyces cerevisiae*
L4	L4p	1-427	4-371	P36578	3U5E_C 4A1A_C	*Saccharomyces cerevisiae* *Tetrahymena thermophila*
L5	L5p	1-178	9-176	P62913	3U5I_J	*Saccharomyces cerevisiae*
L6	L6p	1-192	1-191	P32969	3U5E_H	*Saccharomyces cerevisiae*
L6e	L6e	1-288	27-288	Q02878	3U5E_E 4A1B_E	*Saccharomyces cerevisiae* *Tetrahymena thermophila*
L8e	L7ae	1-266	21-266	P62424	3U5E_G	*Saccharomyces cerevisiae*
L10	L10p	1-317	5-284	P05388	3U5I_q 3A1Y_G 3JSY_A	*Saccharomyces cerevisiae* *Pyrococcus horikoshii* *Methanococcus jannaschii*
L11	L11p	1-165	1-163	P30050	3U5I_K	*Saccharomyces cerevisiae*
L13	L13p	1-203	2-203	P40429	3U5E_O	*Saccharomyces cerevisiae*
L13e	L13e	1-211	2-211	P26373	4A1B_U	*Saccharomyces cerevisiae*
L14	L14p	1-140	8-140	P62829	3U5I_V	*Saccharomyces cerevisiae*
L14e	L14e	1-215	1-139	P50914	3U5I_M	*Saccharomyces cerevisiae*
L15	L15p	1-148	2-148	P46776	3U5E_a	*Saccharomyces cerevisiae*
L15e	L15e	1-204	2-204	P61313	3U5I_N	*Saccharomyces cerevisiae*
L16	L10e	1-214	2-214	Q96L21	3U5E_I	*Saccharomyces cerevisiae*
L18	L18p	1-297	9-297	P46777	3U5E_D	*Saccharomyces cerevisiae*
L18e	L18e	1-188	1-188	Q07020	3U5I_Q	*Saccharomyces cerevisiae*
L19e	L19e	1-196	1-189	P84098	3U5E_R	*Saccharomyces cerevisiae*
L20e	L18ae	1-176	2-176	Q02543	3U5E_S	*Saccharomyces cerevisiae*
L21e	L21e	1-160	2-160	P46778	3U5I_T	*Saccharomyces cerevisiae*
L22	L22p	1-184	2-153	P18621	3U5E_P	*Saccharomyces cerevisiae*
L22e	L22e	1-128	15-126	Q6P5R6	3U5E_U	*Saccharomyces cerevisiae*
L23	L23p	1-156	36-156	P62750	3U5E_X	*Saccharomyces cerevisiae*
L24	L24p	1-145	2-134	P61254	3U5E_Y	*Saccharomyces cerevisiae*
L24e	L24e	1-157	1-124	P83731	3U5I_W	*Saccharomyces cerevisiae*
L27e	L27e	1-136	2-136	P61353	3U5E_Z	*Saccharomyces cerevisiae*
L28e	L28e	1-137	1-137	P46779	4A1B_O	*Tetrahymena thermophila*
L29	L29p	1-123	1-123	P42766	4A1A_U	*Tetrahymena thermophila*
L29e	L29e	1-159	2-79	P47914	3U5E_b	*Saccharomyces cerevisiae*
L30	L30p	1-248	20-248	P18124	3U5I_F	*Saccharomyces cerevisiae*
L30e	L30e	1-115	10-109	P62888	4A1B_G	*Tetrahymena thermophila*
L31e	L31e	1-125	12-124	P62899	4A1A_W	*Tetrahymena thermophila*
L32e	L32e	1-135	1-133	P62910	3U5I_e	*Saccharomyces cerevisiae*
L33e	L35ae	1-110	2-110	P18077	3U5E_f	*Saccharomyces cerevisiae*
L34e	L34e	1-117	2-115	P49207	3U5E_g	*Saccharomyces cerevisiae*
L36e	L36e	1-105	1-103	Q9Y3U8	3U5I_i	*Saccharomyces cerevisiae*
L37e	L37e	1-97	2-91	P61927	3U5I_j	*Saccharomyces cerevisiae*
L38e	L38e	1-70	2-70	P63173	3U5E_k	*Saccharomyces cerevisiae*
L39e	L39e	1-51	2-51	P62891	3U5E_l	*Saccharomyces cerevisiae*
L40e	L40e	77-128	77-128	P62987	3U5I_m	*Saccharomyces cerevisiae*
L41e	L41e	1-25	1-25	P62945	3U5E_n	*Saccharomyces cerevisiae*
L43e	L37ae	1-92	3-92	P61513	3U5I_p	*Saccharomyces cerevisiae*
L44e	L44e	1-106	2-106	P83881	3U5E_o	*Saccharomyces cerevisiae*
P1	L12	1-114	7-63	P05386	3A1Y_D & F	*Pyrococcus horikoshii*
P2	L12	1-115	1-56	P05387	3A1Y_C & E	*Pyrococcus horikoshii*

Supplementary Table 5 | Small subunit, eEF2 and Vig2 protein models from *D. melanogaster*.

Protein name	Protein family	Protein length	Modeled sequence	Uniprot_ID	Template used	Template organism
RACK1	RACK1	1-318	1-318	O18640	3U5C_g	*Saccharomyces cerevisiae*
S1e	S3ae	1-268	17-236	P55830	3U5G_B	*Saccharomyces cerevisiae*
S2	S2p	1-313	6-223	P38979	3U5G_A	*Saccharomyces cerevisiae*
S3	S3p	1-246	3-229	Q06559	3U5C_D	*Saccharomyces cerevisiae*
S4	S4p	1-195	4-184	P55935	3U5G_J	*Saccharomyces cerevisiae*
S4e	S4e	1-261	1-261	P41042	3U5G_E	*Saccharomyces cerevisiae*
S5	S5p	1-267	38-264	P31009	3U5G_C	*Saccharomyces cerevisiae*
S6e	S6e	1-248	1-231	P29327	3U5C_G	*Saccharomyces cerevisiae*
S7	S7p	1-228	39-228	Q24186	3U5G_F	*Saccharomyces cerevisiae*
S7e	S7e	1-194	1-194	Q9VA91	3U5C_H	*Saccharomyces cerevisiae*
S8	S8p	1-130	2-130	P48149	3U5C_W	*Saccharomyces cerevisiae*
S8e	S8e	1-208	2-208	Q8MLY8	3U5C_I	*Saccharomyces cerevisiae*
S9	S9p	1-148	1-148	Q9W237	3U5C_Q	*Saccharomyces cerevisiae*
S10	S10p	1-120	18-119	P55828	3U5G_U	*Saccharomyces cerevisiae*
S10e	S10e	1-163	1-95	Q9VB14	3U5C_K	*Saccharomyces cerevisiae*
S11	S11p	1-151	18-151	P14130	3U5C_O	*Saccharomyces cerevisiae*
S12	S12p	1-143	1-143	Q8T3U2	3U5G_X	*Saccharomyces cerevisiae*
S12e	S12e	1-139	21-139	P80455	3U5C_M	*Saccharomyces cerevisiae*
S13	S13p	1-152	6-142	P41094	3U5G_S	*Saccharomyces cerevisiae*
S14	S14p	1-56	5-56	Q9VH69	3U5C_d	*Saccharomyces cerevisiae*
S15	S15p	1-151	2-151	Q03334	3U5C_N	*Saccharomyces cerevisiae*
S17	S17p	1-155	1-155	Q0E9B6	3U5C_L	*Saccharomyces cerevisiae*
S17e	S17e	1-131	1-120	P17704	3U5G_R	*Saccharomyces cerevisiae*
S19	S19p	1-148	8-131	Q7JZW2	3U5G_P	*Saccharomyces cerevisiae*
S19e	S19e	1-156	1-154	P39018	3U5G_T	*Saccharomyces cerevisiae*
S21e	S21e	1-83	1-82	O76927	3U5G_V	*Saccharomyces cerevisiae*
S24e	S24e	1-131	4-129	Q9W229	3U5C_Y	*Saccharomyces cerevisiae*
S25e	S25e	1-117	40-113	P48588	3U5G_Z	*Saccharomyces cerevisiae*
S26e	S26e	1-114	2-108	P13008	3U5G_a	*Saccharomyces cerevisiae*
S27e	S27e	1-84	1-84	Q9VBU9	3U5C_b	*Saccharomyces cerevisiae*
S28e	S28e	1-65	4-65	Q9W334	3U5G_c	*Saccharomyces cerevisiae*
S30e	S30e	73-132	74-131	Q9VDH8	3U5C_e	*Saccharomyces cerevisiae*
S31e	S27ae	77-156	77-156	P15357	3U5C_f	*Saccharomyces cerevisiae*
eEF2	eEF2	1-844	4-49, 54-844	P13060	1U2R_A 2XQD_Z 2NPF_T	*Saccharomyces cerevisiae* *Saccharomyces cerevisiae* *Thermus thermophilus*
Vig2	Vig2	1-412	108-148, 212-228	Q960D3	3U5C_h	*Saccharomyces cerevisiae*

Supplementary Table 6 | Large subunit proteins from *D. melanogaster*.

Protein name	Protein family	Protein length	Modeled sequence	Uniprot_ID	Template used	Template organism
L1	L1p	1-217	1-217	Q9VTP4	2HW8_A	*Thermus thermophilus*
L2	L2p	1-256	1-253	Q9V3G1	3U5E_A	*Saccharomyces cerevisiae*
L3	L3p	1-416	1-414	O16797	3U5E_B	*Saccharomyces cerevisiae*
L4	L4p	1-401	2-393	P09180	3U5E_C 4A1A_C	*Saccharomyces cerevisiae* *Tetrahymena thermophila*
L5	L5p	1-184	1-182	P46222	3U5I_J	*Saccharomyces cerevisiae*
L6	L6p	1-190	1-190	P50882	3U5E_H	*Saccharomyces cerevisiae*
L6e	L6e	1-243	16-243	Q9V9W2	3U5E_E 4A1B_E	*Saccharomyces cerevisiae* *Tetrahymena thermophila*
L8e	L7ae	1-271	31-271	P46223	3U5E_G	*Saccharomyces cerevisiae*
L10	L10p	1-317	5-227	P19889	3U5I_q 3A1Y_G 3JSY_A	*Saccharomyces cerevisiae* *Pyrococcus horikoshii* *Methanococcus jannaschii*
L11	L11p	1-165	6-163	Q9W1B9	3U5I_K	*Saccharomyces cerevisiae*
L13	L13p	1-205	1-205	Q9VNE9	3U5E_O	*Saccharomyces cerevisiae*
L13e	L13e	1-218	2-211	P41126	4A1B_U	*Saccharomyces cerevisiae*
L14	L14p	1-140	7-140	P48159	3U5I_V	*Saccharomyces cerevisiae*
L14e	L14e	1-166	1-159	P55841	3U5I_M	*Saccharomyces cerevisiae*
L15	L15p	1-149	1-149	P41092	3U5E_a	*Saccharomyces cerevisiae*
L15e	L15e	1-204	2-204	O17445	3U5I_N	*Saccharomyces cerevisiae*
L16	L10e	1-218	2-218	O61231	3U5E_I	*Saccharomyces cerevisiae*
L18	L18p	1-299	8-297	Q9W5R8	3U5E_D	*Saccharomyces cerevisiae*
L18e	L18e	1-188	2-188	Q9VS34	3U5I_Q	*Saccharomyces cerevisiae*
L19e	L19e	1-203	1-203	P36241	3U5E_R	*Saccharomyces cerevisiae*
L20e	L18ae	1-177	5-177	P41093	3U5E_S	*Saccharomyces cerevisiae*
L21e	L21e	1-159	2-159	Q9V9M7	3U5I_T	*Saccharomyces cerevisiae*
L22	L22p	1-186	2-186	Q9W3W8	3U5E_P	*Saccharomyces cerevisiae*
L22e	L22e	1-299	184-299	P50887	3U5E_U	*Saccharomyces cerevisiae*
L23	L23p	1-277	158-277	Q9W0A8	3U5E_X	*Saccharomyces cerevisiae*
L24	L24p	1-149	2-132	Q9VVU2	3U5E_Y	*Saccharomyces cerevisiae*
L24e	L24e	1-155	1-130	Q9VJY6	3U5I_W	*Saccharomyces cerevisiae*
L27e	L27e	1-135	2-135	Q9VBN5	3U5E_Z	*Saccharomyces cerevisiae*
L28e	L28e	1-144	2-141	Q9VZS5	4A1B_O	*Tetrahymena thermophila*
L29	L29p	1-123	1-123	Q9W499	4A1A_U	*Tetrahymena thermophila*
L29e	L29e	1-76	2-76	Q24154	3U5E_b	*Saccharomyces cerevisiae*
L30	L30p	1-252	24-252	P32100	3U5I_F	*Saccharomyces cerevisiae*
L30e	L30e	1-111	10-109	Q9VJ19	4A1B_G	*Tetrahymena thermophila*
L31e	L31e	1-124	14-124	Q9V597	4A1A_W	*Tetrahymena thermophila*
L32e	L32e	1-134	1-132	P04359	3U5I_e	*Saccharomyces cerevisiae*
L33e	L35ae	1-157	1-157	Q9VNB9	3U5E_f	*Saccharomyces cerevisiae*
L34e	L34e	1-162	1-113	Q9VBH8	3U5E_g	*Saccharomyces cerevisiae*
L36e	L36e	1-115	1-113	P49630	3U5I_i	*Saccharomyces cerevisiae*
L37e	L37e	1-93	2-93	Q9VXX8	3U5I_j	*Saccharomyces cerevisiae*
L38e	L38e	1-70	1-70	Q9W5N2	3U5E_k	*Saccharomyces cerevisiae*
L39e	L39e	1-51	2-51	O16130	3U5E_l	*Saccharomyces cerevisiae*
L40e	L40e	77-128	77-128	P18101	3U5I_m	*Saccharomyces cerevisiae*
L41e	L41e	1-25	1-25	Q962S2	3U5E_n	*Saccharomyces cerevisiae*
L43e	L37ae	1-92	2-92	Q9VMU4	3U5I_p	*Saccharomyces cerevisiae*
L44e	L44e	1-104	1-104	Q9VLT7	3U5E_o	*Saccharomyces cerevisiae*
P1	L12	1-112	-	P08570	-	-
P2	L12	1-113	-	P05389	-	-

Supplementary Table 7 | Small subunit rRNA expansion segment definitions, extended nomenclature, length and GC content.

Escherichia coli					*Saccharomyces cerevisiae*			*Drosophila melanogaster*			*Homo sapiens*		
Coordinates	Length (nts)	Variable region	Location (helix)	Expansion segment	Coordinates	Length (nts)	GC (%)	Coordinates	Length (nts)	GC (%)	Coordinates	Length (nts)	GC (%)
16S 63-104	42	1	h6	(-)	18S 59-88	30	(-)	18S 59-87	29	(-)	18S 58-87	30	(-)
16S 140-146	7	2	h7, h8	ES2S	18S 125-142	18	33	18S 124-139	16	31	18S 124-144	21	62
16S 179-218	40	3	h9, h10	ES3S	18S 176-288	113	36	18S 173-293	121	40	18S 178-336	159	72
16S 260-266	7	4	h11	(-)	(-)[a]	0	(-)	(-)[a]	0	(-)	(-)[a]	0	(-)
16S 403-498	96	5	h16, h17	(-)	18S 474-544	71	(-)	18S 479-552	74	(-)	18S 522-593	72	(-)
16S 592-650	59	6	h21	ES6S[b]	18S 639-860	222	44	18S 647-946	300	35	18S 688-917	230	61
16S 840-846	7	7	h26	ES7S	18S 1051-1067	17	29	18S 1138-1154	17	35	18S 1108-1124	17	53
16S 992-1046	55	8	h33	(-)	18S 1217-1266	50	(-)	18S 1305-1353	49	(-)	18S 1274-1323	50	(-)
16S 1120-1153	34	9	h39	ES9S	18S 1340-1384	45	38	18S 1427-1577	151	36	18S 1397-1448	52	60
16S 1256-1257	2	10	h41	ES10S	18S 1489-1493	5	20	18S 1681-1685	5	0	18S 1551-1557	7	57
16S 1321-1322	2	11	h42	(-)	18S 1558-1559	2	(-)	18S 1750-1751	2	(-)	18S 1622-1623	2	(-)
16S 1446-1456	11	12	h44	ES12S	18S 1682-1719	38	55	18S 1874-1914	41	54	18S 1746-1788	43	84
16S 250	1	13	h11	ES13S[c]	18S 319-322	4	50	18S 324-327	4	25	18S 367-370	4	50
16S 876	1	14	h25	ES14S	18S 1096-1100	5	40	18S 1183-1188	6	17	18S 1153-1157	5	40
16S 1284-1286	3	15[d]	h41	(-)	18S 1521-1523	3	(-)	18S 1713-1715	3	(-)	18S 1585-1587	3	(-)

Abbreviations used: expansion segment (ES), rRNA helix of the small subunit (h), nucleotides (nts); ES are highlighted in blue.

[a] Structure is identical to *E. coli*. VR is included in the table since it has been defined in[27].
[b] Includes ES6A, ES6B, ES6C as defined in[10] and h21.
[c] Named ES4B in[10].
[d] Located within variable region 10 according to[27].

Extended expansion segment definition and nomenclature

Variable regions (VR) were defined by comparing the rRNA structures of *E. coli*, *S. cerevisiae*, *D. melanogaster* and *H. sapiens* with numbering according to[27]. A threshold of three structurally differing nucleotides was chosen, meaning that differences in one or two sequential nucleotides are not considered a VR. Previously not defined VRs were added with consecutive numbering (VR13 to VR15 for the small subunit (Supplementary Table 7), VR42 to VR58 for the large ribosomal subunit (Supplementary Table S8)). VRs that are expanded by at least three nucleotides in one of the eukaryotic species in comparison to the bacterial rRNA are defined as an ES. ES numbering is derived from the corresponding VR number followed by "S" or "L" for ESs of the small or large ribosomal subunit, respectively. This results in novel ES13S, ES14S, ES43L, ES44L and ES45L for the human and *Drosophila* rRNA.

Supplementary Table 8 | Large subunit rRNA expansion segment definitions, extended nomenclature, length and GC content.

Escherichia coli					*Saccharomyces cerevisiae*			*Drosophila melanogaster*			*Homo sapiens*		
Coordinates	Length (nts)	Variable region	Location (helix)	Expansion segment	Coordinates	Length (nts)	GC (%)	Coordinates	Length (nts)	GC (%)	Coordinates	Length (nts)	GC (%)
23S 1-14	14	1	H1	(-)	5.8S 1-3	3	(-)	5.8S 1-2	2	(-)	5.8S 1-3	3	(-)
23S 84-102	19	2	H7	(-)	5.8S 71-87	17	(-)	5.8S 70-85	16	(-)	5.8S 71-87	17	(-)
23S 137-142	6	3	H9	ES3L	5.8S 122-131	10	40	5.8S 119-123, 2S 1-4	9	56	5.8S 120-130	11	73
23S 159-167	9	4	H10	ES4L	5.8S 148-158, 25S 1-10	21	48	2S 21-30, 28S 1-14	24	38	5.8S 147-157, 28S 1-9	20	70
23S 271-272	2	5	H15	ES5L	25S 115-160	46	43	28S 119-168	50	42	28S 114-163	50	78
23S 365-368	4	6	H16	(-)	25S 261-265	5	(-)	28S 279-283	5	(-)	28S 272-276	5	(-)
23S 538-555	18	7	H25	ES7L	25S 430-629	200	49	28S 449-779	331	24	28S 441-1306	866	83
23S 607-621	15	8	H28	ES8L	25S 681-699	19	37	28S 831-849	19	26	28S 1358-1381	24	75
23S 638-655	18	9	H30, H31	ES9L	25S 716-786	71	52	28S 866-986	121	30	28S 1398-1503	106	80
23S 845-851	7	10	H38	ES10L	25S 977-986	10	40	28S 1177-1199	23	4	28S 1694-1726	33	73
23S 876-901	26	11	H38	(-)	25S 1012-1038	27	(-)	28S 1225-1251	27	(-)	28S 1752-1778	27	(-)
23S 926-933	8	12	H38	ES12L	25S 1063-1104	42	36	28S 1276-1317	42	29	28S 1803-1842	40	75
23S 1022-1026	5	13	H41, H42	ES13L	25S 1191-1201	11	55	28S 1405-1416	12	42	28S 1930-1940	11	64
23S 1150-1151[a]	0	14	H41	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 1170-1179	10	15	H45	ES15L	25S 1345-1359	15	47	28S 1560-1600	41	20	28S 2084-2272	189	85
23S 1204-1206	3	16	H46	(-)	25S 1384-1386	3	(-)	28S 1625-1627	3	(-)	28S 2298-2300	3	(-)
23S 1223-1226	4	17	H46	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 1282-1288	7	18	H48	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 1373	1	19	H52	ES19L	25S 1554-1582	29	48	28S 1797-1813, 28S 1859-1866[c]	25	16	28S 2468-2506	39	74
23S 1410-1424	13	20	H54, H55	ES20L	25S 1619-1653	35	40	28S 1903-1965	63	44	28S 2543-2597	55	73
23S 1451-1460	10	21	H57	(-)	25S 1681-1687	7	(-)	28S 1993-1999	7	(-)	28S 2625-2631	7	(-)
23S 1482-1509	28	22	H58	(-)	25S 1708-1736	29	(-)	28S 2021-2049	29	(-)	28S 2653-2681	29	(-)
23S 1525-1528	4	23	H58, H59	(-)	25S 1753-1756	4	(-)	28S 2067-2070	4	(-)	28S 2698-2701	4	(-)
23S 1532-1539	8	24	H59	(-)	25S 1760-1765	6	(-)	28S 2074-2078	5	(-)	28S 2705-2711	7	(-)
23S 1543-1546	4	25	H56, H59	(-)	25S 1769-1772	4	(-)	28S 2082-2085	4	(-)	28S 2715-2718	4	(-)
23S 1576-1592	15	26	H54, H55	ES26L	25S 1804-1825	22	36	28S 2117-2141	25	52	28S 2750-2773	24	62
23S 1713-1745	33	27	H63	ES27L	25S 1945-2103	159	57	28S 2261-2482	222	32	28S 2894-3607	714	87
23S 1857-1885	29	28	H68	(-)	25S 2217-2228	12	(-)	28S 2595-2606	12	(-)	28S 3721-3732	12	(-)
23S 2091-2092	2	29	H75, H76	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 2129-2159	31	30	H78	ES30L	25S 2471-2174	4	50	28S 2849-2868	20	40	28S 3975-4036	62	92
23S 2203-2220	18	31	H79	ES31L	25S 2519-2588	70	50	28S 2913-3120	208	29	28S 4081-4165	85	85
23S 2286-2287	2	32	H82, H83	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 2296-2322	27	33	H84	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 2396-2397	2	34	H88	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)

Suppl. Table 8 (continued) | Large subunit rRNA expansion segment definitions, extended nomenclature, length and GC content.

Escherichia coli					*Saccharomyces cerevisiae*			*Drosophila melanogaster*			*Homo sapiens*		
Coordinates	Length (nts)	Variable region	Location (helix)	Expansion segment	Coordinates	Length (nts)	GC (%)	Coordinates	Length (nts)	GC (%)	Coordinates	Length (nts)	GC (%)
23S 2405-2412	8	35	H88	(-)	25S 2776-2784	9	(-)	28S 3308-3316	9	(-)	28S 4353-4361	9	(-)
23S 2627-2629	3	36	H73, H94	(-)	25S 2996	1	(-)	28S 3527-3531	5	(-)	28S 4573-4574	2	(-)
23S 2674-2675	2	37	H95, H96	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 2702-2705	4	38	H96	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)	(-)[b]	0	(-)
23S 2789-2810	22	39	H98	ES39L	25S 3152-3294	143	51	28S 3686-3868	183	33	28S 4729-4966	238	82
23S 2832-2835	4	40	H100, H101	(-)	25S 3316-3320	5	(-)	28S 3890-3894	5	(-)	28S 4988-4992	5	(-)
23S 2856-2861	6	41	H101	ES41L	25S 3341-3363	23	39	28S 3915-3937	23	52	28S 5013-5035	23	65
23S 123-128	6	42	H8	(-)	5.8S 109-113	5	(-)	5.8S 106-110	5	(-)	5.8S 108-111	4	(-)
23S 276-294	19	43	H16-H18	ES43L	25S 164-183	20	55	28S 172-197	26	27	28S 167-192	26	88
23S 316-317[b]	0	44	H19, H20	ES44L	(-)[b]	0	(-)	28S 222-227	6	17	28S 216-219	4	75
23S 344-361	18	45	H16-H18	ES45L	25S 233-257	25	48	28S 255-275	21	29	28S 248-268	21	100
23 383-391	9	46	H21	(-)	25S 280-285	6	(-)	28S 298-303	6	(-)	28S 291-296	6	(-)
23S 411-416	6	47	H22	(-)	25S 304-311	8	(-)	28S 322-329	8	(-)	28S 315-322	8	(-)
23S 526-532	7	48	H2, H25	(-)	25S 420-424	5	(-)	28S 438-443	6	(-)	28S 431-435	5	(-)
23S 1271-1274	4	49	H26, H47	(-)	25S 1452-1455	4	(-)	28S 1694-1497	4	(-)	28S 2366-2369	4	(-)
23S 1807-1810	4	50	H66	(-)	25S 2165-2169	5	(-)	28S 2544-2547	4	(-)	28S 3669-3673	5	(-)
23S 1846-1849	4	51	H68	(-)	25S 2205-2209	5	(-)	28S 2583-2587	5	(-)	28S 3709-3713	5	(-)
23S 2643-2645	3	52	H94, H95	(-)	25S 3010-3012	3	(-)	28S 3545-3547	3	(-)	28S 4588-4590	3	(-)
23S 2732-2735	4	53	H96, H97	(-)	25S 3099-3101	3	(-)	28S 3633-3635	3	(-)	28S 4677-4679	3	(-)
23S 2769-2770	2	54	H97, H94	(-)	25S 3133-3135	3	(-)	28S 3667-3669	3	(-)	28S 4711-4713	3	(-)
23S 2776-2781	6	55	H94	(-)	25S 3141-3144	4	(-)	28S 3675-3678	4	(-)	28S 4719-4721	3	(-)
23S 2883-2903	21	56	H1, H99	(-)	25S 3387-3396	10	(-)	28S 3961-3970	10	(-)	28S 5059-5070	12	(-)
5S 73-76	4	57	Loop E	(-)	5S 71-77	7	(-)	5S 70-75	6	(-)	5S 70-75	6	(-)
5S 87-89	3	58	Loop D	(-)	5S 89-92	4	(-)	5S 87-90	4	(-)	5S 87-90	4	(-)

Abbreviations used: expansion segment (ES), rRNA helix of the large subunit (H), nucleotides (nts); ES are highlighted in blue.

[a] Numbers indicate position within the 23S rRNA. Variable region is absent in *E. coli*.
[b] Structure is identical to *E. coli*. VR is included in the table since it has been defined in[27].
[c] *Drosophila* ES19L contains 45 nts (1814-1858) that are cleaved out of the mature 28S rRNA.

Supplementary Table 9 | Contacts of eEF2 and the human 80S ribosome.

eEF2 (residue number)	Interaction partner: residue range	eEF2 domain
27-29	28S: 4604-4607	I
66, 68	28S: 4607-4608	
109	28S: 4605-4606	
132, 162	L6: 98	
136-138	28S: 4605-4606	
159	28S: 4601	
162	28S: 4599, 4601	
166-169	L6: 96, 98-100	
166	L6: 116	
170, 173	L6: 141	
179-180	L10: 139-140, 142, 144, 146	
183	L10: 135, 139, 145-147	
190-191	L10: 132, 147, 150	
186-187	L10: 132, 135, 147	
197-198, 202	L10: 149	
204	L10: 147, 149-150	
261-262	L10: 146	G’
263	L10: 155	
406-409, 526	18S: 478-480	II
448, 460	18S: 488	
526	18S: 480	III
552	18S: 4607	
517-518, 521	S12: 50	
515	S12: 73	
518-519	S12: 75	
517, 521, 538, 544	S12: 95-97	
596-597, 720	28S: 3760	IV
625	S30e: 11	
629, 633, 647	18S: 1319-1322	
655-656	S30e: 7	
667	S12: 84	
667-670, 673	18S: 1503-1506	
670-671	18S: 1328-1330	
677	S30e: 8	
677	18S: 615	
684, 680-681	S30e: 5, 7-8	
685	S30e: 3-5	
712	18S: 1332	
671, 710-714, 716	SERPB1:, 162-165, 168	
710-712, 715	SERPB1: 154-155	
719-720	18S: 1826	
727, 853	28S: 3762	
753, 782	L11: 31	V
756-757, 760	28S: 2008-2009	
761, 764-765, 803	28S: 4419	
760, 764, 769-771, 773	28S: 1981-1982	
772, 774-779	L11: 25-27, 29-41	
778-779	L11: 34	
779	L10: 132-133	
798, 806	28S: 4605	
801	28S: 4477, 4605-4606	

Supplementary Table 10 | Contacts of eEF2 and the *Drosophila* 80S ribosome.

eEF2 (residue number)	Interaction partner: residue range	eEF2 domain
28-29	28S: 3558, 3561, 3563-3564	I
66, 68	28S: 3564-3565	
112-113	28S: 3562-3564	
136, 166	L6: 96	
136, 171, 173	L40e: 77	
140, 114	28S: 3562- 3563	
166	28S: 3558-3559	
170-172	L6: 94, 96-98, 100	
171-174, 177	L6: 139	
180, 183-184	L10: 136, 139-140, 144	
187-188, 190-191,194-195	L10: 132-136, 139, 145-147, 149-150	
197-200	L10: 149	
207	L10: 147-150	
209	L10: 146	
253	L10: 144, 156, 158	G'
267	L10: 142, 144, 158	
268	L10: 162	
273	L10: 142	
394	18S: 436	II
417, 418, 420, 427, 429	S12: 142	
425	S12: 139	
429, 464-466	S12: 141, 142-143	
435	18S: 445	
504	S12: 50	III
507, 508, 524	S12: 97, 99	
511	S12: 143	
512	18S: 437	
537-538, 541	28S: 3563-3564	
582, 584	28S: 2632-2634	IV
583	18S: 1952	
611	S30e: 86-87	
615, 619, 633	18S: 1349-1351	
630-631, 634	S31e: 77-78	
651-655	S31e: 77-80	
653-654, 656-657, 659, 662	18S: 1633-1634, 1639, 1357-1359	
657, 660, 696-700, 702	Vig2: 120-121, 124, 127-131	
663	18S: 573	
663, 666-667	S30e: 80	
670	S30e: 79-80, 83	
671	S30e: 75-77	
674	S30e: 77, 79	
691-693	S31e: 79	
698	18S: 1361	
700, 705-706	18S: 1951	
709, 713, 831, 839	28S: 2634-2637	
712	S30e: 75	
739, 742-744, 746, 750, 759, 768	28S: 1484-1486	V
740	L10: 134	
746, 749-750, 756-758	28S: 1457-1458	
750-751	28S: 3374	
758, 760	L11: 25-27	
761, 763-765	L11: 30-31	
762, 763	L11: 34, 40	
765-767	L10: 130-132	
784, 787-788, 792	28S: 3562-3563	
788	L40e: 128	

Supplementary Table 11 | Contacts of SERBP1 and the human 80S ribosome.

SERPB1 (residue number)	Interaction partner: residue range	Position on the small subunit
139, 141-144	18S: 1236-1238	mRNA tunnel
139, 142-143	18S: 1523-1525	
140	E-tRNA: 27-28	
143-145	E-tRNA: 29-30, 32, 35	
149,151	18S: 1246-1248	
152, 153-155, 158-159	18S: 132	
154-155	eEF2: 710-712, 715	
157-161	18S: 1699-1702	
159-160, 162-163	18S: 1490	
161-163	18S: 1330-1332	
162, 164, 166-172, 177	S3: 145-150	
162, 165, 168	eEF2: 671, 710-714, 716	
165	S12: 64	
165-167	18S: 614-615, 626	
167-170	18S: 1699	
170	S5: 120	
174-175	18S: 624-625, 629-631	
175-177	S3: 138-143	
176-180, 183, 186	S3: 114-117	
180, 183	S5: 124	
180, 183-184	S5: 147	
181	S5: 109, 111	
183, 186-187	S3: 120-121,	
185, 187-188	S30e: 51-53	
186	18S: 607	
187	S3: 124	
188	S5: 151	
281-287	S10e: 83, 85-86, 88-91, 94	Small subunit head
286-287, 290	S10e: 16-18	
293	S12e: 28, 116	
293	S10e: 86	
300	S12e: 48	
302-303	S12e: 44-46	
303	S31e: 127	

Supplementary Table 12 | Contacts of Vig2 and the *Drosophila* 80S ribosome.

Vig2 (residue number)	Interaction partner: residue range	Position on the small subunit
108-110	18S: 1266-1268, 1652	mRNA tunnel
109, 112-113	E-tRNA: 28, 30, 32, 34-35	
117-119	18S: 1276-1278	
120-121, 124, 127-131	eEF2: 657, 660, 696-700, 702	
121, 124-125, 127-130	18S: 1361-1362	
123-127	18S: 1829-1830	
125-126, 128-130	18S: 1619	
126, 129-138, 132-133, 135-138	S3: 146-150	
126, 132, 134-138	18S: 1825-1827	
131-133	18S: 585	
132	18S: 1630	
136	S5: 103	
138	18S: 11	
138	S5: 96	
139	18S: 583	
141, 142, 144	18S: 566, 587, 589-590	
140, 142-143	S3: 144-145, 150, 152, 154	
143-146	S3: 117-118, 140	
146-148	S5: 92-94, 105-107, 130	
148	18S: 560	
212-216, 219	S10e: 86-89	Small subunit head
216-217, 220	S10e: 13, 16-17	
216, 219-220, 223	S10e: 81, 83-84	
212-220	S10e: 74,78, 83-91	
223	S12e: 31, 34	
223, 226, 228	S12e: 120, 122-123	

Supplementary Table 13 | Ribosomal RNA GC content.

Saccharomyces cerevisiae			*Drosophila melanogaster*			*Homo sapiens*		
rRNA	Length (nts)	GC (%)	rRNA	Length (nts)	GC (%)	rRNA	Length (nts)	GC (%)
40S ES	467	41	40S ES	661	37	40S ES	538	65
40S core	1333	46	40S core	1334	45	40S core	1331	52
40S total	1800	45	40S total	1995	43	40S total	1869	56
60S ES	975	49	60S ES	1494	30	60S ES	2641	83
60S core	2700	48	60S core	2704	46	60S core	2707	55
60S total	3675	48	60 total	4198	40	60S total	5348	69
80S ES	1442	46	80S ES	2155	32	80S ES	3179	80
80S core	4033	47	80S core	4038	46	80S core	4038	54
80S total	5475	47	80S total	6193	41	80S total	7217	66

Abbreviations: expansion segments (ES), nucleotides (nts).

Supplementary Information References

72. Gonzalez, I. L. et al. Variation among human 28S ribosomal RNA genes. *Proc Natl Acad Sci U S A* **82**, 7666-7670 (1985).

73. Ware, V. C., Renkawitz, R. & Gerbi, S. A. rRNA processing: removal of only nineteen bases at the gap between 28S alpha and 28S beta rRNAs in Sciara coprophila. *Nucleic Acids Res* **13**, 3581-3597 (1985).

74. Marton, M. J., deAldana, C. R. V., Qiu, H. F., Chakraburtty, K. & Hinnebusch, A. G. Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosomes in activation of eIF2 alpha kinase GCN2. *Molecular and Cellular Biology* **17**, 4474-4489 (1997).

75. Seidelt, B. et al. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412-1415 (2009).

76. Armache, J. P. et al. Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Res* **41**, 1284-1293 (2013).